USING TEST OF INTRANSITIVITY TO COMPARE COMPETING STATIC AND DYNAMIC MODELS OF INTERTEMPORAL CHOICE

Junyi Dai

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements for the degree

Doctor of Philosophy

in the Department of Psychological and Brain Sciences,

Indiana University

July 2014

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

Jerome Busemeyer, Ph.D.

John Kruschke

Peter Todd

James Townsend

May 5, 2014

To Xiaoqing Tu, my beloved wife,

who sacrificed so much along my journal toward the degree.

Acknowledgements

I would like to thank my advisor Dr. Busemeyer, who continuously provides intellectual and financial support for my academic career. Without his support, my current success is unimaginable.

I would like to thank Dr. Kruschke, Dr. Todd, and Dr. Townsend, for their wonderful service on my committee.

I would like to thank Prof. Le from East China Normal University, who initiated my attempt to study abroad.

Finally, I would like to thank my parents for their continuous solicitude and initial financial support without which my study in America is impossible.

Junyi Dai

USING TEST OF INTRANSITIVITY TO COMPARE COMPETING

STATIC AND DYNAMIC MODELS OF INTERTEMPORAL CHOICE

Most traditional research on intertemporal choice assumes a deterministic, static, and alternative-wise perspective, leading to the widely adopted delay discounting paradigm. Recently, however, Dai and Busemeyer (2014) demonstrated that intertemporal choice is probabilistic, dynamic, and attribute-wise in nature, and they developed an attribute-wise diffusion model to account for these properties. This dissertation advances the previous research. Specifically, two new experiments with different types of intertemporal choice questions and an even more extensive comparison of competing static and dynamic models were conducted to further examine the relevant properties and look for a more comprehensive cognitive model of intertemporal choice. The results of the first experiment indicated that the probabilistic, dynamic, and attribute-wise nature of intertemporal choice was supported under both conditions when the SS options occurred immediately or were delayed options. In addition, the results of the second experiment indicated that most participants showed transitive intertemporal preferences in terms of weak stochastic transitivity. The extensive model comparison led to an overall best model which was a generalization of the diffusion model with direct differences as advocated in Dai and Busemeyer. This model can account for all the effects and phenomena examined in this dissertation, including the delay duration effect, the common difference effect (and its reversal), the magnitude effect, and the potential intransitivity of intertemporal choice, as well as the marginal and conditional relationships between choice proportions and response times observed in individual data as a demonstration of the dynamic nature of intertemporal choice. Furthermore, this model can be conveniently extended to intertemporal choice between losses and account for

the relevant gain-loss asymmetry. Consequently, it is recommended as a replacement for the existing models of intertemporal choice which assume a deterministic, static, and alternative-wise perspective on the topic.

Table of Contents

Intertemporal choice involves tradeoffs between benefits and costs at different times (Frederick, Loewenstein, & O'Donoghue, 2002). Examples of intertemporal choice are quite common in the economic world as well as in our everyday lives. For instance, a decision to deposit a portion of one's income into a bank account rather than to spend the money right now can be interpreted as an intertemporal choice. The two options involved in this example are, on the one hand, using the money to fulfill some of one's immediate need, and on the other hand, saving the money to get more money and satisfaction in the future. Broadly speaking, any decision with a time element can be viewed as an intertemporal choice, such as a decision to continue one's education after acquiring a bachelor's degree instead of starting a formal job right after graduation from college. This decision will certainly bring about multiple outcomes to be realized at different time points in future, and thus constitutes a good example of intertemporal choice.

Participants of behavioral studies on intertemporal choice are usually required to show their preference between two options with different payoffs and delays. For example, researchers may ask participants to choose between receiving 10 dollars in 10 days and receiving 20 dollars in 30 days on a specific trial. Oftentimes the first option is referred to as a smaller-but-sooner (SS) option, and the second is referred to as a larger-but-later (LL) option. By collecting choice data for various pairs of intertemporal options, researchers hope to find mechanisms governing intertemporal choice and then characterize each individual accordingly.

Most previous research on intertemporal choice adopted a deterministic, static, and alternative-wise approach to the topic, leading to the extremely popular delay discounting paradigm for intertemporal choice. According to this paradigm, when facing an intertemporal choice, people would first calculate the discounted utility of each option in a certain way and then choose for sure the option with a higher discounted utility. Nonetheless, a number of recent studies have demonstrated some critical deficits of the delay discounting paradigm and thus imposed a severe challenge to the deterministic, static, and alternative-wise approach to intertemporal choice. For instance, the results of a sequence of studies by Scholten, Read, and colleagues (e.g., Roelofsma & Read, 2000; Scholten & Read, 2010; Scholten, Read, & Sanborn, 2014) suggested that people actually process intertemporal choice in an attribute-wise rather than alternative-wise manner. An attribute-wise manner means that people first compare the two options on each attribute (i.e., money or delay attribute) and then combine the results to make a decision, whereas an alternative-wise manner implies that people first calculate the utility of each option separately and then make a choice by comparing their utilities as suggested by the delay discounting paradigm. More recently, Dai and Busemeyer (2014) investigated the probabilistic and dynamic nature of intertemporal choice and compared a large number of relevant stochastic models to find a substitute for traditional models built upon the delay discounting paradigm. The results of relevant studies strongly supported a probabilistic and dynamic perspective on intertemporal choice as opposed to the traditional deterministic and static view. In addition, the winning model with an attribute-wise view not only described both choice and response time data well but also offered a comprehensive framework to explain three important effects in intertemporal choice, i.e., the delay duration effect, common difference effect, and magnitude effect.

Although recent progress in the research area clearly advocates a probabilistic, dynamic, and attribute-wise approach to intertemporal choice, there are many important additional issues to be addressed before we can confidently adopt the new approach. Consequently, two new experiments were conducted for this dissertation to provide more supportive information for the new perspective. In addition, an even more extensive model comparison was performed to further compare competing static and dynamic models of intertemporal choice. The ultimate goal of this dissertation is to obtain a more comprehensive understanding of the affective and cognitive processes underlying intertemporal choice and to find a better model that accounts for a majority of effects and phenomena in intertemporal choice.

This dissertation is organized as follows. First, I will provide a brief review of intertemporal choice literature with a focus on the traditional delay discounting paradigm. Second, I will elaborate on recent work concerning intertemporal choice that suggests a fundamentally different approach, especially that from Dai and Busemeyer (2014) which advocates not only an attribute-wise view on intertemporal choice but also a probabilistic and dynamic perspective on the topic. Third, I will introduce the various effects and phenomena in intertemporal choice examined in this dissertation. Fourth, I will discuss new issues in the literature that should be addressed before the probabilistic, dynamic, and attribute-wise approach can be widely adopted with confidence. Fifth, I will summarize all the probabilistic models explored in this dissertation. Sixth, I will report the empirical results of Experiment 3 in Dai and Busemeyer and two new experiments designed to address the remaining issues, as well as the preliminary results of model comparison using data from the three studies. An overall model comparison will follow and the prediction of the resulting best model will be shown. The dissertation ends with a general discussion on various implications of the model comparison results and a conclusion in favor of a paradigm shift in intertemporal choice research.

Traditional Approach to Intertemporal Choice

Due to its ubiquitousness and significance in both economic activities and everyday life, intertemporal choice has been a hot research topic in economics, psychology, and decision sciences for a long time. It was first examined by Scottish economist John Rae (1834) who provided an in-depth discussion of the psychological motives underlying intertemporal choice, including the bequest motive, the tendency to exert self-control, the uncertainty of human life, the excitement produced by the expectation of immediate consumption, and the pain of postponing this consumption. Following this line of research, other economists such as William S. Jevons (1888), Herbert S. Jevons (1905), and Eugen von Böhm-Bawerk (1889) proposed additional factors that affect intertemporal choice, including impulsivity, anticipated utility of future consumption, and underestimation of future needs. Böhm-Bawerk's work was later on formalized by American economists Irving Fisher (1930), who proposed yet another factor, fashion, which he believed to have significant impact on intertemporal choice. There is no doubt that early exploration of economists on the topic of intertemporal choice was heavily influenced by psychological considerations that deserve much more attention in the new century if we want to obtain a comprehensive understanding of the issue.

In 1937, Paul Samuelson proposed the discounted utility (DU) model, the most popular theory of intertemporal choice among economists since its birth. This model condenses all the factors economists proposed previously into a single index of discount rate and proposes that the discount rate is constant over time and independent of many characteristics of potential consumption, such as its magnitude and type. On the one hand, this model provides a parsimonious account of intertemporal choice which builds up a convenient platform for further exploration into the topic. On the other hand, the proposal of reducing various intertemporal motives into a single factor and a number of other assumptions of the model might be unrealistic in practice. In fact, Samuelson did not endorse the DU model as a normative model, nor was he confident in its descriptive validity. Nonetheless, its alluring simplicity and elegance soon turned it into the standard tool economists used to analyze intertemporal decisions.

About four decades after the publication of Samuelson's article on DU model, psychologists began to resurrect the psychological orientation on the topic of intertemporal choice by conducting empirical research and proposing different models that provided a better description of the actual data from human and animal subjects. For instance, George Ainslie and colleagues (Ainslie, 1974, 1975; Ainslie & Herrnstein, 1981) studied pigeons' choice behavior between food reinforcements with different delay durations and found that the preference of pigeons was better described by hyperbolic discounting curves than by exponential ones suggested by the DU model. The authors interpreted the results as an issue of impulsivity and impulse control, which clearly echoed the early proposals of economists before the DU model gained dominance in intertemporal choice research. This line of research resulted in the hyperbolic discounting model by Mazur (1987), which set up the foundation for a majority of psychological studies on intertemporal choice afterwards. The fundamental difference between the hyperbolic discounting model and the DU model lies in their predictions on time consistency. Specifically, the DU model suggests that people's preference between two payoffs at different time points should not change when time elapses, whereas the hyperbolic discounting model might predict a preference reversal under the same condition, as was observed in empirical studies on both animal and human subjects (e.g., Ainslie & Herrnstein, 1981; Kirby & Herrnstein, 1995; Thaler, 1981).

In the meantime, a new generation of economists including George Loewenstein, Drazen Prelec, and Richard Thaler put a great deal of effort to explore intertemporal choice in order to modify relevant economic models to account for different types of behavioral anomalies clearly revealed in empirical studies. For example, Loewenstein and Prelec (1992) summarized the anomalies in intertemporal choice with regard to the DU model and provided an interpretation by proposing a behavioral model with different forms of value function and discount function. Similarly, Prelec and Loewenstein (1991) compared intertemporal choice to risky choice and presented a common approach to various effects in these two areas that are similar to each other. Although most of their work was within the framework of economic analysis and theorization, there were definitely certain psychological insights involved in their theories, such as the inclusion of a reference point in the value function. It is no doubt that both economic and psychological research on this topic has contributed to a rich foundation on which new thoughts and theories of intertemporal choice can be developed and tested.

Intertemporal Choice from a Delay Discounting Perspective

Despite a long history of examination and contributions from both economists and psychologists, a great proportion of intertemporal choice research up to now has been governed by what may be an unwarranted perspective on how people actually make such a choice, i.e., a delay discounting perspective. According to this perspective, when encountering a scenario of intertemporal choice, people will first figure out the subjective value (i.e., utility) of each option and then always choose the option with the highest subjective value. Because in most conditions people would like to get a certain payoff immediately rather than at a later time, it seems plausible to assume that the subjective value of a payoff is discounted when delayed. This leads to the pivotal notion of delay discounting in the traditional approach to intertemporal choice. Consequently, finding an appropriate form of the time discount function which determines the ratio of discounted utility of a delayed payoff to its instantaneous utility if occurring immediately becomes the key issue to address in the research area. However, there exist a number of other possible explanations for people's general tendency to prefer an immediate payoff to a delayed one of the same objective amount. For example, people may want to get the payoff right now just because waiting is painful as suggested by early economists investigating intertemporal choice, rather than that the delayed payoff has a lower subjective value as suggested by the delay discounting perspective. Before proceeding to offer different perspectives and accounts for intertemporal choice, let me first discuss a little more about several discount functions widely used in traditional research on the topic. This will be helpful for developing more complicated probabilistic models in later sections.

Exponential discount function. As mentioned above, the DU model assumes a constant discount rate over time. This assumption leads to an exponential form of discount function which reflects the degree to which the subjective value of a delayed payoff should be discounted relative to its subjective value when occurring immediately. Specifically, the exponential discount function can be represented as

$$D(t) = exp(-kt) \tag{1}$$

where t indicates the delay duration of a payoff and k is a discounting parameter, which should assume a positive value to guarantee delay discounting. In addition, the larger the value of k is, the more rapidly the subjective value of a delayed option will be discounted. Let DU represent the discounted utility of an option and u(v) be the utility or subjective value of an immediate payoff with money amount v, Then the DU model entails that a person should prefer an SS option, (v_s, t_s), to an LL option, (v_l, t_l), when

$$d = DL_{LL} - DL_{SS} = u(v_l) \cdot D(t_l) - u(v_s) \cdot D(t_s) = u(v_l) \cdot exp(-kt_l) - u(v_s) \cdot exp(-kt_s) \tag{2}$$

is negative, and prefer the LL option when d is positive. The quantity d is actually the difference in discounted utility between the two intertemporal options and Equation 2 suggests that people would always choose the option with a higher discounted utility. Because the exponential discount function implies a constant discount rate, the ratio of DU_{LL} to DU_{SS} will remain the same when time passes. Consequently, the sign of d will not change over time and therefore the exponential discount function suggests time-consistent preference, which is oftentimes regarded as a requirement in economic analysis as a principle of rationality (Strotz, 1956).

Hyperbolic discount function. In reality, however, people do not always behave as consistently as suggested by economic theories based on abstract axioms. In the area of intertemporal choice it means that people's preference between two payoffs at different times may change when time elapses. Specifically, when both the SS and LL options get closer, the attractiveness of the SS option might loom larger at a higher rate than that of the LL option. Consequently, when two options are far away from now, people may prefer the LL option, but the same people may want to take the SS option instead when time passes. This is exactly what researchers found in both human and animal subjects (e.g., Christensen-Szalanski, 1984; Green, Fisher, Perlow, & Sherman, 1981; Millar & Navarick, 1984). This pattern of preference reversal is typically referred to as the common difference effect and it produces time-inconsistent behavior (Loewenstein & Prelec, 1992). A common response to this type of preference reversal is to adopt a different discount function which can accommodate this anomalous change according to the DU model. There are a number of candidates for the alternative discount function, among which the most popular and simplest one is the following hyperbolic discount function proposed by Mazur (1987),

$$D(t) = \frac{1}{1+kt} \tag{3}$$

where t denotes the delay duration of an option and k is a discounting parameter with higher values indicating more rapid discounting just as in the exponential discount function (i.e., Equation 1). The hyperbolic discount function and the related hyperbolic discounting model suggest that people would always choose the SS option when

$$d = DL_{LL} - DL_{SS} = u(v_l) \cdot D(t_l) - u(v_s) \cdot D(t_s) = \frac{u(v_l)}{1+kt_l} - \frac{u(v_s)}{1+kt_s} \tag{4}$$

is negative, and would always choose the LL option when d is positive. The only difference between Equations 2 and 4 lies in the form of discount function being used. Nonetheless, the hyperbolic discount function is fundamentally different from the exponential discount function in that the former predicts declining rather than constant discount rate over time. Specifically, the per-period discount factor according to the exponential discount function equals $exp[-k(t+1)]/exp(-kt) = exp(-k)$, which is independent of t, the front-end delay. To the contrary, the per-period discount factor according to the hyperbolic discount function equals $\frac{1}{1+k(t+1)} / \frac{1}{1+kt} = \frac{1+kt}{1+k(t+1)} = 1 - \frac{k}{1+k(t+1)}$, which is an increasing function of t. Because there is a one-to-one mapping between discount rate and discount factor, i.e.,

$$\rho = \frac{1}{f} - 1 \tag{5}$$

where ρ is the discount rate and f is the discount factor (Frederick et al., 2002), the exponential discount function predicts a constant discount rate over time while the hyperbolic discount function predicts a declining discount rate. In other words, according to the hyperbolic discount function, the farther a unit-length interval is from now, the less discounting it will produce on the subjective value of a delayed payoff.

Although the hyperbolic discount function can accommodate the preference reversal in intertemporal choice, it is not without problems. It is no surprise that the simplicity of the function form makes it unlikely to provide a sufficient account for the declining of subjective value of payoffs delayed to different extents. In fact, a recent study by Christian Luhmann (2013) found that the declining of discount rate when payoffs are delayed further into the future is more moderate than what is suggested by a hyperbolic discount function. That is to say, the actual tendency of preference reversal is weaker than what is predicted by a hyperbolic discount function. Therefore, it seems necessary to revise the hyperbolic function (i.e., Equation 3) to capture better the regularity in empirical data. A number of more complicated discount functions have been put forward for this purpose. For example, Green, Fry, and Myerson (1994) proposed the following two-parameter hyperboloid function,

$$D(t) = \frac{1}{(1 + kt)^s}, \tag{6}$$

and Rachlin (2006) proposed a two-parameter hyperbola function,

$$D(t) = \frac{1}{1 + kt^s}. \tag{7}$$

Both functions have one more parameter, i.e., the parameter s, than the hyperbolic function, Consequently, they are more general and thus can account for a larger proportion of the variance at both group and individual levels than the hyperbolic function (McKerchar et al., 2009). Nonetheless, the interpretations of the extra parameter differ between the two discount functions. In Equation 6, the parameter s is supposed to reflect the scaling of delay duration and/or money amount that differ among individuals, while in Equation 7, the parameter is presumed to reflect the sensitivity of discounted utility to the delay duration t. The latter was explicitly derived from Stevens' (1957) power law, which connects objective magnitude of a physical stimulus to its subjectively perceived intensity or strength. This suggests the critical role time perception plays in intertemporal choice and I will elaborate on this issue when introducing more general models in a later section.

Alternative Approaches to Intertemporal Choice

Although the delay discounting paradigm has been widely adopted by both economists and psychologists for a long period of time, it is built upon three unquestioned but fundamental background assumptions about how people make decisions in general and make intertemporal choices in particular. First, the delay discounting paradigm assumes implicitly a deterministic view on intertemporal choice. Stated otherwise, it presumes that people's preference between a pair of intertemporal options is fixed and they would always choose the same option when facing the pair of options repeatedly. This is true for the DU model, the hyperbolic discounting model, and its variants. Second, all models derived from the delay discounting paradigm are static in the sense that they are silent on the underlying dynamic processes leading to the explicit choices. Consequently, they do not provide any account for the decision time associated with an intertemporal choice. Third, models based on the delay discounting paradigm assume an alternative-wise approach to intertemporal choice, which is a common feature of all utility-based models. Specifically, such an approach suggests that people evaluate each intertemporal option separately before comparing the (discounted) utilities to make a choice. In summary, the traditional approach to intertemporal choice, i.e., the delay discounting paradigm, assumes a deterministic, static, and alternative-wise view on intertemporal choice. This view, however, might impose unnecessary constraints on intertemporal choice research and thus result in misleading conclusions and implications. Therefore, we need to consider alternative approaches to the topic to gain more valid understanding of intertemporal choice and more effective methods to help people make better intertemporal decisions.

Deterministic versus Probabilistic Approaches

It is a common observation that people change their mind from time to time. Therefore, a probabilistic approach to choice behavior seems to be more realistic than a strictly deterministic approach as assumed by the delay discounting paradigm. In fact, the probabilistic nature of choice behavior has been well documented in the literature of risky choice. For example, Mosteller and Nogee's (1951) classic work revealed inconsistent preferences at an individual level for single gambles over repeated occasions. According to a deterministic perspective on risky choice, when the positive outcome of a mixed gamble increased, the preference function in observed proportion of taking the gamble ought to have a step form with a step from zero to one at the point where a player was indifferent between the mixed gamble and a neutral option. Nonetheless, the actual preference function generated from empirical data closely resembled the smooth, gradually increasing S-shaped psychometric functions commonly found in psychophysical studies. This result suggested that the deterministic perspective is inadequate to provide a reasonably good account for risky choice behavior in human subjects.

Despite its intuitive validity, the probabilistic nature of intertemporal choice had been literally ignored in the literature. Recently, however, Dai and Busemeyer (2014) examined this property with three experiments in which each participant needed to answer a large set of intertemporal choice questions associated with the delay duration effect, the common difference effect, and the magnitude effect (described later). The results, especially that of the third experiment with repeatedly presented questions, provided strong evidence for the probabilistic nature of intertemporal choice. For example, each participant in the last experiment switched between the SS and LL options in multiple questions when they were asked repeatedly. Clearly we need a probabilistic approach to human choice behavior to account for this result.

Static versus Dynamic Approaches

The second assumption of the delay discounting paradigm, i.e, the static view on intertemporal choice, is also debatable. On the one hand, the static view ignores the issue of decision time, making static models of human decision-making easier to construct than dynamic models. On the other hand, decisions differ from one another in terms of the deliberation or decision time required and the details of the underlying processes. More important, deliberation time may change systematically as a function of choice probability (Diederich, 2003; Dai & Busemeyer, 2014) and therefore provides further information to distinguish between competing models. Consequently, a comprehensive approach to intertemporal choice should take decision time into consideration and provide a description of the underlying dynamic processes leading to the explicit choice. A number of dynamic stochastic models of decision-making have been proposed and applied to account for results of empirical studies on preferential choice (e.g., Brown & Heathcote, 2008; Krajbich, Armel, & Rangel, 2010; Usher & McClelland, 2004). Among them, decision field theory (DFT) by Busemeyer and colleagues (Busemeyer & Townsend, 1992, 1993; Johnson & Busemeyer, 2005; Roe, Busemeyer, & Townsend, 2001) has been most widely applied and is the principal dynamic theory on risky choice. Because of the similarity between risky choice and intertemporal choice (Prelec & Loewenstein, 1991), it is natural to extend DFT to the research area of intertemporal choice. Consequently, Dai and Busemeyer (2014) developed and tested a large number of dynamic diffusion models of intertemporal choice, including those based on DFT, and compared them with static models of various types. It turned out that dynamic models in general performed better than static models in describing the choice data, not to mention their advantage when decision time was of concern.

Alternative-wise versus Attribute-wise Approaches

Given the dominant status of the concept of utility in economic analysis, most economic theories assume an alternative-wise perspective on preferential choice. The alternative-wise perspective implies that each option has an inherent utility or subjective value that is independent of other options and thus could be separately evaluated. This idea was also adopted by Samuelson when he developed the DU model and inherited by psychologists who tried to find a better form of the discount function to describe empirical data. When the actual mental processes underlying preferential choice are of concern, however, the notion of independent utility appears to be unwarranted and people might indeed use a quite different way to make such a choice. For example, some computational models of risky choice, such as the proportional difference (PD) model (González-Vallejo, 2002) and the priority heuristic (Brandstätter, Gigerenzer, & Hertwig, 2006), actually suggest that people first compare different alternatives within each attribute and then combine the results to make a decision. The same also applies to DFT models, which assume that within-attribute differences among different alternatives are the building blocks of the diffusion process via which preference for each alternative accumulates over time.

Because the delay discounting paradigm has long dominated intertemporal choice research, attribute-wise models of intertemporal choice are gaining momentum in the literature. Among them, the tradeoff model proposed by Scholten and Read (2010) and the DFT model by Dai and Busemeyer (2014) seem to be the most promising ones. The tradeoff model was developed to account for a variety of anomalies that elude the delay discounting paradigm, such as nonadditive discounting and the inseparability between time and money in intertemporal choice. It also provides an explanation for intransitivity of intertemporal choice, which seems to be supported by empirical data (Roelofsma & Read, 2000). The DFT model, by contrast, provides a dynamic account of the processes underlying intertemporal choice and thus provides a description of the decision time associated with each choice. A direct comparison of these two models was conducted in this dissertation for a better model of intertemporal choice.

Effects and Phenomena in Intertemporal Choice

In addition to the aforementioned time-inconsistent behavior and the associated common difference effect found in empirical studies, researchers have also discovered various other effects and phenomena in intertemporal choice, such as the magnitude effect (Kirby & Marakoviæ, 1996; Green, Myerson, & McFadden, 1997; Chapman & Winquist, 1998), the delay duration effect (Dai & Busemeyer, 2014), and the potential intransitivity of intertemporal choice (Roelofsma & Read, 2000). The magnitude effect indicates that the discounting parameter in Equation 3 is a decreasing function of the size of a delayed reward, and therefore larger amounts of money tend to suffer less from delay discounting than smaller ones (Loewenstein & Prelec, 1992). The delay duration effect indicates that people's preference between two delayed options will shift toward the SS option when the delay durations of the two payoffs are increased proportionally (i.e., multiplied by a common constant greater than one). Both effects can be explained by a single assumption within the common framework for risky and intertemporal choice proposed by Prelec and Loewenstein (1991). According to the assumption of increasing proportional sensitivity, increasing all (positive) values of an attribute by a common multiplicative constant makes the attribute more influential in a decision. Take the magnitude effect as an example. Suppose that a person has no preference between receiving 10 dollars now and receiving 20 dollars in a month. Then when both reward amounts are multiplied by 2 to generate a new pair of options, i.e., ($20, now) versus ($40, 1 month), the reward amount will become more influential according to the assumption of increasing proportional sensitivity. Consequently, the person should turn to prefer the option with a higher value on reward amount, i.e., the LL option. When interpreting this change of preference under the framework of hyperbolic discounting , it is readily seen that the larger dollar amount (i.e., \$40 in the second pair) suffer less from delay discounting and thus is associated with a smaller discounting parameter than the smaller dollar amount (i.e., \$20 in the first pair).

Another central issue of choice behavior in general and intertemporal choice in particular is the (in)transitivity of preference. A deterministic interpretation of preference transitivity entails that for any triplet of options $\{A, B, C\}$, if A is preferred to or indifferent from B and B is preferred to or indifferent from C by an individual, then the very person should prefer A to C or be indifferent between them. Symbolically, this version of preference transitivity can be represented as if $A \succeq B$ and $B \succeq C$, then $A \succeq C$. Here, the symbol $\succeq$ reads as "is preferred to or indifferent from". When the probabilistic nature of choice behavior is taken into account, the deterministic interpretation of preference transitivity should be transformed into a stochastic version. Several possible ways have been proposed for this purpose, including weak stochastic transitivity (WST; Davidson & Marschak, 1959), moderate stochastic transitivity (MST; Grether & Plott, 1979), and strong stochastic transitivity (SST; Tversky & Russo, 1969). WST implies that, for a triplet of options $\{A, B, C\}$, if $Pr(A|\{A, B\}) \geq .5, Pr(B|\{B, C\}) \geq .5$, then $Pr(A|\{A, C\})$ should also be greater than or equal to .5. Similarly, MST requires that $Pr(A|\{A, C\}) \geq min(Pr(A|\{A, B\}), Pr(B|\{B, C\}))$ and SST requires that $Pr(A|\{A, C\}) \geq max(Pr(A|\{A, B\}), Pr(B|\{B, C\}))$ given the same condition as in WST. Clearly, WST imposes the weakest constraints among the three stochastic models of transitivity (Cavagnaro & Davis-Stober, in press).

The significance of preference transitivity in choice behavior comes form the fact that it is a prerequisite for all types of alternative-wise, utility-based models, such as the expected utility theory (von Neumann & Morgenstern, 1944) for choice under risk and the DU model (Samuelson, 1937) for intertemporal choice. Consequently, examination of preference transitivity could provide critical information for discovering mechanisms underlying choice behavior and thus facilitate model selection. As a result, Roelofsma and

Read (2000) examined this issue in the context of intertemporal choice. In their study, each participant was required to make six binary choices between all possible pairs of four options similar to those used in the classic work of Tversky (1969) concerning preference intransitivity in risky choice. Specifically, the four intertemporal options involved were (monetary amounts were in Dutch Guilders): (a) (7, 1 week) (b) (8, 2 weeks) (c) (9, 4 weeks) (d) (10, 7 weeks) The authors categorized individual choices into three patterns, transitive (denoted as T), intransitive with a circle among three options (denoted as I_3), and intransitive with a circle among four options (denoted as I_4). It turned out that more than half of the participants (46 out of 88) showed intransitive choice pattern, and the ratio of I_4 patterns to I_3 patterns in the empirical data across participants was much higher than what would be expected from two different probabilistic models of intertemporal choice (i.e., the constant error model and Thurstone's Case V model to be discussed in a later section) built upon the exponential discount function (i.e., Equation 1). Based on these results, the authors suggested that intertemporal preference might be intransitive and that intertemporal choice was made through an attribute-wise decision strategy. Consequently, they proposed a version of Tversky's (1969) lexicographic semiorder rule to explain the intransitive choice patterns.

At the first glance, the empirical results and relevant analysis of Roelofsma and Read (2000) seemed to provide evidence for intransitivity of intertemporal preference. The research and analysis method used in the article, however, might be inadequate to establish the very property. For instance, because participants answered each of the six distinct questions only once, it was impossible to establish intransitive intertemporal preference at an individual level when choice variability (i.e., choosing different options when the same pair of options is presented repeatedly) was taken into account. Furthermore, their group-level analysis which considered the issue of choice variability took into account only two possible ways to generate probabilistic models of intertemporal choice. As will become clear later in this dissertation, there are more classes of probabilistic models we can apply to intertemporal choice. Finally, it is possible to find intransitive patterns in an aggregate analysis even if individual intertemporal preferences are completely transitive in a deterministic way. Consequently, in this dissertation participants were required to answer each intertemporal choice question in their stimulus set multiple times and more probabilistic models of intertemporal choice were proposed and tested. In this way, we could investigate the issue of transitivity of intertemporal preference at an individual level and use the results to select a best model among a large group of candidate models.

The issue of intransitivity in intertemporal preference is also related to the phenomena of superadditivity and subadditivity in delay discounting. Superadditivity in delay discounting arises when the time interval between options counts more if it is taken as a whole than if it is divided into shorter subintervals, whereas subadditivity in delay discounting suggests the opposite pattern (Scholten & Read, 2010). Both nonadditive patterns in delay discounting are inconsistent with traditional models such as the DU model and hyperbolic discounting model. This is due to the fact that both of them suggest delay duration of each option rather than the interval between options determines the degree of discounting. In extreme cases, superadditivity and subadditivity can lead to intransitive preferences in intertemporal choice. For example, a person may prefer ($12, 2 weeks) to ($10, 1 week) and prefer ($14, 3 weeks) to ($12, 2 weeks) but prefer ($10 dollars, 1 week) to ($14 dollars, 3 week). In this case, the time interval between the two options in the last pair (i.e., 2 weeks) produces more discounting when taken as a whole than when divided into subintervals (i.e., the two intervals involved in the first two pairs of options). Read and colleagues (e.g., Read, 2001; Read & Roelofsma, 2003) have conducted a number of studies to demonstrate the superadditivity and subadditivity in delay discounting. As the study on intransitivity of intertemporal preference, these studies either did not consider choice variability in individual responses or used an aggregate analysis. Therefore, the interpretation of the results may not be valid at an individual level when considering choice variability. As stated above, this dissertation adopts a different approach to examining intransitivity of intertemporal preference. This will shed new light on the issue of nonadditivity in delay discounting as well.

New Important Issues

Although Dai and Busemeyer (2014) reported empirical results and results of model comparison favoring a probabilistic, dynamic, and attribute-wise approach to intertemporal choice, there are several additional important issues to address before we can confidently advocate a shift of research paradigm in intertemporal choice. First, almost all the questions in Dai and Busemeyer's (2014) studies involved delayed SS and LL options as opposed to immediate SS options and delayed LL options. It is possible that people would adopt an attribute-wise strategy in the former case but an alternative-wise strategy in the latter. This is because in the latter case one only needs to calculated the discounted utility of the delayed LL options but not that of the SS options. By contrast, when both options are delayed ones, an alternative-wise strategy involves calculating two discounted utilities in each question. This will make the alternative-wise strategy more demanding and thus less likely to be adopted. Furthermore, most traditional studies on intertemporal choice actually used the latter form of questions to study people's time preference and distinguish between different discount functions. Therefore, before calling for a fundamental change in research paradigm, it is necessary to show that the probabilistic, dynamic, and attribute-wise nature of intertemporal choice also holds for choice questions between immediate SS options and delayed LL options.

Second, the three studies reported in Dai and Busemeyer (2014) were all designed around the delay duration effect, the common difference effect, and the magnitude effect in intertemporal choice and did not examine the important issue of preference transitivity. As mentioned above, preference transitivity is a necessary condition for any alternative-wise, utility-based model of choice behavior. Although Roelofsma and Read (2000) provided evidence for intransitive intertemporal choice and presented an attribute-wise model as a solution, the problems in their experimental design and analysis method appeared to weaken the validity of their conclusions. Consequently, it is important to conduct a new study in which individuals answer the same questions repeatedly so that choice variability can be taken into consideration and individual level analysis can be performed. If intransitive preference is revealed again at an individual level, then we can have more confidence in the attribute-wise approach to intertemporal choice.

Third, Dai and Busemeyer (2014) demonstrated the marginal relationship between choice proportions and response times in all three studies but left for future research the conditional relationship between choice proportions and response times. The marginal relationship suggests that questions with extreme choice proportions (i.e., close to 0 or 1) tend to have shorter response times than questions with moderate choice proportions (i.e., around .5), whereas the conditional relationship means that within each question, the option with a higher choice probability (i.e., above .5) tends to have a shorter response time than the other option. The former is consistent with diffusion models in general while the latter favors a specific form of diffusion model with initial bias contingent on mean drift rate (or equivalently with different thresholds for the two responses). Therefore, examining both relationships in empirical data will be conducive to model development and selection. This will also help clarify the dynamic nature of intertemporal choice.

Probabilistic Models of Intertemporal Choice

Because intertemporal choice is essentially probabilistic as shown in Dai and Busemeyer (2014), it is necessary to build relevant models that can account for the choice variability in empirical data. To this end, Dai and Busemeyer developed a large number of stochastic models of intertemporal choice and fit them to the choice data. In addition, various dynamic diffusion models were also fit to both choice and response time data simultaneously so that more information from the data set could be employed to select the best model. In this dissertation an even more extensive set of models, including two probabilistic versions of the tradeoff model, are fit and compared to find a model that can account for all the empirical results found in the reported studies. In what follows, I provide a detailed description of all the models explored in this dissertation. Most of the models can be categorized in terms of the way they transform objective value and time into subjective ones, their core theories, and their stochastic specifications which make them capable of predicting probabilistic choice patterns.

Transformations of Objective Value and Time

Ever since the early days of economic analysis on human choice behavior, researchers have realized that the subjective value, or utility, of a payoff is not identical or linear to its objective amount. For example, the famous example of St. Petersburg paradox (Bernoulli, 1954) suggests that people do not maximize the expected value when deciding whether to take a gamble or not. Instead, they consider the utility of a payoff, which represents the amount of pleasure or usefulness it produces. The utility, or subjective value, of a payoff usually does not increase linearly with its objective value but at a declining speed. This leads to the popular concept of diminishing marginal utility which means that each extra unit of money generates less and less utility. A variety of utility functions have been proposed to capture this property, among which the power utility function (i.e., $u(v) = v^{\alpha}$) involved in prospect theory (Kahneman & Tversky, 1979) and cumulative prospect theory (Tversky & Kahneman, 1992) appears to be the most popular substitute for an identity utility function. To the contrary, the tradeoff model (Scholten & Read, 2010) adopts a different form of utility function for gains as follows:[1]

$$u(v) = \frac{1}{\gamma} log(1 + \gamma v). \tag{8}$$

The authors chose this specific form to fulfill several qualitative requirements of intertemporal choice, such as the absolute magnitude effect and the proportional magnitude effect. The former is the same as the magnitude effect discussed above, while the latter suggests that the absolute compensation $v_l - v_s$ is larger for a larger v_s to remain indifference between a pair of options with fixed t_s and t_l. Because this form of utility function, together with other components of the tradeoff model, could account for a large number of empirical results in intertemporal choice research, it is worth a try in this dissertation.

The nonlinear relationship between objective and subjective times, however, was not well recognized and incorporated into intertemporal choice research until recently. For example, Zauberman, Kyu Kim, Malkoc, and Bettman (2009) found that participants' subjective perception of prospective duration was nonlinear and concave in objective time, and that subjective time perception could be altered by priming, leading to a reduction of hyperbolic discounting. Similarly, Takahashi, Oono, and Radford (2008) combined psychophysical functions of time perception with both exponential and hyperbolic discount functions to generate new forms of discount function and found that Weber-Fechner function of time perception and exponential discount function combined provided the best account for both aggregate and individual data. These and other studies (e.g., Wittmann & Paulus, 2008) suggest that time perception plays a critical role in intertemporal choice and incorporating time transformation functions into extant models may fundamentally change our understanding of this topic. Consequently, in this dissertation I explore a number of time transformation functions, some of which were reported in Dai and Busemeyer and the others were not. Specifically, four classes of time transformation functions are examined here, including an identity function (i.e., $p(t) = t$), a power function (i.e., $p(t) = t^{\beta}$), the time weighing function involved in the tradeoff model (i.e., $p(t) = \frac{1}{\tau} log(1 + \tau t)$), and a power function with an extra scaling constant (i.e., $p(t) = ct^{\beta}$). The motive for the last class of functions is to put subjective value and time on comparable dimensions so that they can be additively combined to generate meaningful overall evaluations of options. Without the scaling constant, the impact of one attribute may be overestimated relative to the other and the measurement unit in either dimension may substantially affect the performance of relevant models. Overall, six pairs of utility function and time transformation function are explored in this dissertation:

$$u(v) = v, p(t) = t \tag{9}$$

$$u(v) = v^{\alpha}, p(t) = t^{\beta}, \alpha \leq 1, \beta \leq 1 \tag{10}$$

$$u(v) = v^{\alpha}, p(t) = ct^{\beta}, \alpha \leq 1, \beta \leq 1, c > 0 \tag{11}$$

$$u(v) = v^{\alpha}, p(t) = t^{\beta}, \alpha \leq 2, \beta \leq 2 \tag{12}$$

$$u(v) = v^{\alpha}, p(t) = t^{\beta}, \alpha \leq 2, \beta \leq 2, c > 0 \tag{13}$$

$$u(v) = \frac{1}{\gamma} log(1 + \gamma v), p(t) = \frac{1}{\tau} log(1 + \tau t), \gamma > 0, \tau > 0 \tag{14}$$

The first pair of transformation functions simply assume identity forms which entail constant marginal utility and marginal subjective time. The next two pairs of transformation functions involve power functions with exponents no larger than one. These functions in general predict diminishing marginal utility and marginal subjective time except when $\alpha = 1$ and/or $\beta = 1$, under which condition the transformations become linear and thus predict constant marginal utility and marginal subjective time as the first pair of transformation functions do. To the contrary, the fourth and fifth pairs of transformation functions allow for exponents greater than one. Consequently, they can predict diminishing, constant, or increasing marginal utility and marginal subjective time. That is, they are more general than the first three pairs of transformation functions. The rationale for relaxing the constraints on the range of exponents is that for both money and time, especially small amounts of money and time, there might be certain cutoff points that have unusual impact on intertemporal choice. For example, a payment of 100 dollars may exert an impact more than twice as much as that of a payment of 50 dollars if a person needs exactly 100 dollars to fulfill a specific goal. Similarly, a delay of 1 year and 1 day might appear much more undesirable than a delay of 364 days because the former is longer than 1 year which is commonly used as a standard duration for planning. Finally, the last pair of transformation functions come from the tradeoff model and they also in general predict diminishing marginal utility and marginal subjective time. It is worth noting that the transformation on objective time can also be interpreted as the (dis)utility of waiting time and thus acquires the same nature as the utility of a payoff.

Core Theories

The core theory of a probabilistic model is the deterministic special case of the model (Loomes & Sugden, 1995). Dai and Busemeyer (2014) used four core theories to generate various probabilistic models of intertemporal choice, and so do I in this dissertation. Specifically, the time transformation functions listed above were incorporated into the DU model (i.e., Equation 2) and the hyperbolic discounting model (i.e., Equation 4) to generate two alternative-wise core theories, and an attention shift mechanism was invoked as the foundation of two attribute-wise core theories. According to the alternative-wise core theory based on the exponential discount function (i.e., Equation 1), one should prefer the LL option if

$$d = DU_{LL} - DU_{SS} = u(v_l) \cdot exp(-kp(t_l)) - u(v_s) \cdot exp(-kp(t_s)) \tag{15}$$

is positive, and prefer the SS option when d is negative. Similarly, the alternative-wise core theory based on the hyperbolic discount function (i.e., Equation 3) suggests that one should choose the LL option if

$$d = DU_{LL} - DU_{SS} = u(v_l)/(1 + kp(t_l)) - u(v_s)/(1 + kp(t_s)) \tag{16}$$

is positive and choose the SS option when d is negative. Probabilistic models derived from these two core theories will be referred to as generalized DU models and generalized hyperbolic models respectively. Note that the traditional DU model (i.e., Equation 2) and hyperbolic discounting model (i.e., Equation 4) are special cases of the generalized models with identity time transformation functions. Furthermore, the two-parameter hyperboloid function proposed by Green et al., (1994) is a special case of the generalized hyperbolic model with power utility function and identity time transformation function, and the two-parameter hyperbola function by Rachlin (2006) is a special case of the generalized hyperbolic model with identity utility function and power time transformation function. Obviously Equations 15 and 16 provide a general framework for most delay discounting models in the literature.

Given the similarity between risky and intertemporal choice (Prelec & Loewenstein, 1991) and the success of DFT models in explaining a variety of empirical findings in risky choice research, Dai and Busemeyer (2014) developed two attribute-wise core theories of intertemporal choice based on DFT. Specifically, they assumed that, when facing an intertemporal choice, people switch their attention between the money and delay attributes and sample the difference within these two attributes. Because the LL option is advantageous on the money attribute over the SS option but disadvantageous on the delay attribute, the deterministic attribute-wise models based on the attention shift mechanism suggest that people will choose the LL option if the weighted difference in subjective value is larger than the weighted difference in subjective time, and that the weight on either attribute is determined by the amount of attention on the attribute. For the same reason, people should choose the SS option if the weighted difference in subjective time is larger than the weighted difference in subjective value.

The two attribute-wise core theories differ in the way within-attribute differences are evaluated. For the core theory based on direct differences, the within-attribute differences are the simple differences in subjective value or time between the two options. That is, the direct difference on the money attribute equals $u(v_l) - u(v_s)$, and the direct difference on the delay attribute equals $p(t_l) - p(t_s)$. To the contrary, the attribute-wise core theory based on relative differences evaluates within-attribute differences in a proportional way. Specifically, the relative (or proportional) difference on the money attribute equals $\frac{u(v_l)-u(v_s)}{u(v_s)}$, and the relative difference on the delay attribute equals $\frac{p(t_l)-p(t_s)}{p(t_s)}$. The relative differences can be viewed as the result of normalizing the direct differences by the local minimum on either attribute. This idea of normalization was borrowed from the PD model (González-Vallejo, 2002), which also employs within-attribute proportional differences to generate probabilistic choice models. However, the normalizers in the PD model are the local maxima instead of minima. I will elaborate on the difference between the PD model and the DFT models later on when contrasting all the probabilistic models of intertemporal choice explored in this dissertation. Mathematically, the attribute-wise core theory with direct differences suggests that people should choose the LL option if

$$d = w \cdot (u(v_l) - u(v_s)) - (1 - w) \cdot (p(t_l) - p(t_s)) \tag{17}$$

is greater than zero and choose the SS option when d is smaller than zero. Similarly, the attribute-wise core theory with relative differences entails that people should choose the LL option if

$$d = w \cdot \frac{u(v_l) - u(v_s)}{u(v_s)} - (1 - w) \cdot \frac{p(t_l) - p(t_s)}{p(t_s)} \tag{18}$$

is positive and choose the SS option if d is negative. The parameters w and $1 - w$ in the above equations represent the attention weights on the money and delay attributes respectively.

Note that when $t_s = 0$ and $w \neq 1$ (i.e., with an immediate SS option and at least some attention on the delay attribute), the attribute-wise core theory with relative differences produces a negatively infinite d value. This in turn leads to deterministic prediction on choice response even when the core theory is coupled with any of the stochastic specifications discussed below to generate probabilistic models. In reality, however, people usually do not always choose an immediate SS option against a delayed LL option. Therefore, it is necessary to modify the core theory for questions with immediate SS options to accommodate the empirical result. To this end, an extra parameter, t_z, is added to the core theory to represent the effective delay of a presumably immediate option. Because any "immediate" reward will be fulfilled at least after participants answer all the questions, it is actually associated with a certain amount of delay. The parameter is constrained between 0 and 1 (in unit of day) to cover a meaningful range of the effective delay duration. If all the questions of concern involve delayed SS options, then the extra parameter is simply ignored in the model fitting and comparison procedure described later.

Stochastic Specifications

When a deterministic core theory is enhanced by a stochastic specification, it becomes capable of making probabilistic predictions on choice behavior and thus accommodating choice variability in empirical data. A large number of probabilistic choice models have been proposed for risky choice and Dai and Busemeyer (2014) expanded the literature by developing similar models for intertemporal choice. Specifically, the authors explored four broad classes of probabilistic models, including constant error models, Fechner models, random utility models, and diffusion models. The first three classes of models are static and thus only capable of predicting choice probability. To the contrary, the diffusion models have a dynamic structure and therefore can be used to fit both choice and response time data. In this section I will describe each of these model classes and how they could be utilized to model intertemporal choice. I will also introduce a couple of other possible models for intertemporal choice, such as the PD model (González-Vallejo, 2002) and the tradeoff model (Scholten & Read, 2010).

Constant error models. The constant error model is the simplest possible probabilistic model of choice behavior. It assumes that people have a true preference between options but might choose the undesirable option with a fixed probability due to "trembling hand" (Harless & Camerer, 1994). Let ϵ represent the chance of "hand trembling" on each trial, then the probability of choosing the preferred option $1-\epsilon$ and the probability of choosing the undesirable one is just ϵ. The value of ϵ is typically assumed to be no larger than 0.5. When a person is indifferent between two options, the constant error model simply assigns equal choice probabilities, i.e., 0.5, to both options.

When intertemporal choice is of concern, we can incorporate the hand trembling mechanism into each of the four core theories aforementioned (i.e., Equations 15-18) to generate corresponding constant error models. We can interpret the quantity d in Equations 15-18 as the difference in true utility as suggested by the original formulation of constant error models. This is a natural interpretation for the alternative-wise core theories (i.e., Equations 15 and 16), but may seem unwarranted for the attribute-wise models (i.e., Equations 17 and 18) since the concept of utility is essentially abandoned in the latter case. However, we can reorganize Equations 17 and 18 to generate pseudo-utility-based models. For example, Equation 17 can be rewritten as

$$d = [w \cdot u(v_l) - (1-w) \cdot p(t_l)] - [w \cdot u(v_s) - (1-w)p(t_s)]. \tag{19}$$

Now we can interpret d as the difference in "utility" between the two options and the original attribute-wise model is turned into an alternative-wise one. This suggests that an alternative-wise model can mimic an attribute-wise model in terms of prediction on choice behavior with a deterministic perspective. The same is also true for certain types of probabilistic models, such as the constant error model and Probit model with fixed σ discussed below. However, the mimicry breaks down with other types of stochastic specification, such as that assumed by the winning model in Dai and Busemeyer (2014). If the previous best model or a similar one still wins in the current research, then the issue of model mimicry should not be a major concern.

Another feature of constant error model is that it predicts an abrupt change in choice probability given certain experimental manipulation that monotonically alters the d value in the core theories. Stated otherwise, a constant error model is similar to a deterministic model in the sense that both are inconsistent with a gradual change pattern in choice proportion. Dai and Busemeyer (2014) actually found gradual changes in choice proportion across multiple experiments when systematically changing the delay durations or reward amounts as required by the delay duration effect and the magnitude effect. Consequently, it was no surprise that constant error models performed poorly in fitting the empirical data.

Logistic models. A similar but more complicated class of probabilistic choice models is the Fechner models (Becker, DeGroot, & Marschak, 1963). Like constant error models, the Fechner models assume that people do have true preference between options. However, when making a choice, they are susceptible to processing errors (Loomes & Sugden, 1995). This processing error varies from trial to trial and may more than offset the true difference in utility between options, leading to a choice of the less preferred option. Accordingly, the probability of choosing option A from a pair of options $\{A, B\}$ equals

$$P(A|\{A, B\}) = Pr(u_A - u_B + \epsilon > 0) \tag{20}$$

where u_A and u_B represent the true utilities of the two options and ϵ denotes the random amount of processing error. Given a proper distribution of the random variable ϵ, we can derive the choice probability of either option from Equation 20. A common practice in this case is to assign a logistic distribution to the error term, leading to the following logistic model,

$$P(A|\{A,B\}) = \frac{1}{1 + exp(-g(u_A - u_B))} = \frac{1}{1 + exp(-g \cdot d)} \tag{21}$$

where $d = u_A - u_B$ represents the true difference in utility between the two options and g is a free parameter. When g is zero, the logistic model becomes a random choice model according to which either option has a choice probability of 0.5. When g approaches infinity, the logistic model reduces into a deterministic model in which the option with a higher true utility will always been chosen.

When considering intertemporal choice, we can combine the stochastic specification of logistic models (i.e., Equation 21) with the core theories on intertemporal choice discussed above (i.e., Equations 15 - 18) to develop probabilistic models. As before, we can rewrite Equations 17 and 18 so that the quantity d can be interpreted as the difference in utility as assumed in the original form of the Fechner model. Like constant error models based on Equations 17 and 18, the corresponding logistic models can also be interpreted from an alternative-wise perspective and these two interpretations cannot be distinguished from each other in terms of prediction on choice probabilities. Logistic models, however, predict a gradual rather than abrupt change pattern in choice probability when d increases monotonically. This prediction is fundamentally different from that of deterministic models and constant error models and thus makes logistic models more competitive in fitting empirical data.

Probit models. Another widely used class of probabilistic choice models is the random utility models. We can generate random utility models by introducing random components into deterministic utility models (Becker et al., 1963). The pivotal difference between deterministic and random utility models lies in the way of assigning and interpreting the utility of each option. A deterministic interpretation of utility suggests that each option is equipped with a certain utility that is fixed across repeated trials, while a random interpretation means that the utility of each option can vary from trial to trial. As a result, the preference relation between options based on the random utilities can also change from trial to trial, leading to observed choice variability. Both types of utility models assume that people would always choose the option with the higher utility at a given time. Mathematically, the choice probability of option A from a pair of options $\{A, B\}$ equals

$$P(A|\{A,B\}) = Pr(U_A > U_B) \tag{22}$$

in which U_A and U_B are the random utilities of the two options respectively. [2]

One common way to turn Equation 22 into a practical model of choice probability is to specify the joint distribution of the random utilities. A typical assumption in this case is that the random utilities follow a bivariate normal distribution with independent components. Consequently, Equation 22 turns into the following Probit model

$$P(A|\{A,B\}) = \Phi(\frac{d}{\sigma}) \tag{23}$$

where Φ denotes the cumulative distribution function of a standard normal distribution and d represents the difference in mean utility between the two options.

Two classes of models can be developed from Equation 23. The first one assumes further that all options have the same utility variability and thus the differences in random utility also have the same variance across pairs of options. The resultant random utility model is actually a Thurstone Case V model (Thurstone, 1927). According to this model, $U_A - U_B \sim N(d, \sigma^2)$ in which $d = u_A - u_B$ represents the difference in mean utility as in Equation 23 and σ is the common standard deviation of the difference in random utility across different pairs of options.

To the contrary, the second class of random utility models with Equation 23 as its stochastic specification presumes that utility variability varies across options. Consequently, the variance of utility difference may change from pair to pair. In this case, the value of σ in Equation 23 differs across different pairs of options and thus the resultant models are more complicated than the first class of models. Dai and Busemeyer (2014) tried different ways of assigning σ values between alternative-wise and attribute-wise models of intertemporal choice. Specifically, for alternative-wise models, they assumed that the longer a payoff is delayed, the more uncertain its random utility will be. The underlying psychological intuition is that the longer a payoff is delayed, the more difficult it is to evaluate its utility due to various factors such as the uncertainty of human life. Mathematically, they set

$$\sigma = \sqrt{c \cdot (p(t_s) + p(t_l))} \tag{24}$$

to reflect this property. In Equation 24, the parameter c is a free parameter to be estimated from actual data.

For attribute-wise models, Dai and Busemeyer (2014) assumed that the attention weight on each attribute equals the probability of sampling the difference on the attribute at any time. This sampling process produces variability in instantaneous evaluation from which one can derive a standard deviation equal to

$$\sigma = \sqrt{w(1-w)}|(u(v_l) - u(v_s)) - (p(t_l) - p(t_s))| \tag{25}$$

for models with direct differences, and

$$\sigma = \sqrt{w(1-w)}|\frac{u(v_l) - u(v_s)}{u(v_s)} - \frac{p(t_l) - p(t_s)}{p(t_s)}| \tag{26}$$

for models with relative differences. With Equations 15 - 18 for core theories and Equations 23 - 26 for stochastic specifications, a large number of random utility models of intertemporal choice were generated and tested in Dai and Busemeyer (2014).

Note that for the second class of random utility models, the choice variability parameter, i.e., σ, is also dependent on the attribute values of the intertemporal options in question. Consequently, the resultant attribute-wise Probit models with parameter σ

specified by Equations 25 and 26 could not be interpreted as an alternative-wise model. This is because the parameter σ is derived from the variability of instantaneous evaluation based on the attention shift assumption, which is essentially attribute-wise. As a result, if one of the attribute-wise Probit models with σ contingent on attribute values wins the model comparison, then we can only interpret the result as a piece of evidence in favor of an attribute-wise decision strategy for intertemporal choice.

Random preference models. Yet another way to generate random utility models of choice behavior is to introduce variability into the principal parameter of the corresponding core theories, such as the k parameter in the alternative-wise core theories or the w parameter in the attribute-wise core theories. In this way, the principal parameter is no longer fixed across time but varies from trial to trial. It assumes a particular value at any given instance, resulting in a certain utility of each option and thus a specific preferential relationship between each pair of options.[3] Due to the variability in the principal parameter, the preference between each pair of options varies across trials and thus the explicit choice. Consequently, this type of random utility models are usually referred to as random preference models (Loomes & Sugden, 1995).

Dai and Busemeyer (2014) generated random preference models of intertemporal choice as follows. For the generalized discounted utility models, they assumed that $exp(-k)$ followed a truncated normal distribution between 0 and 1; for the generalized hyperbolic discounting models, they assumed that $log(k)$ followed a normal distribution; and for the attribute-wise models, they assumed that the attention weight parameter, i.e., w, followed a truncated normal distribution between 0 and 1. We can interpret these models as if the discount rate or the attention weights on money and delay attributes for each person varies across time and such variation leads to explicit variability in choice behavior. The range of $exp(-k)$ for the generalized discounted utility models was chosen to fulfill the requirement of delay discounting, and the range of the attention weight parameter for the attribute-wise models was set to cover the whole spectrum of possible normalized amount of attention to either attribute. The distributional form of $log(k)$ in the generalized hyperbolic discounting models was chosen based on the results of previous research which suggested that the natural logarithm of k parameter in the hyperbolic discount function (i.e., Equation 3) follows a normal distribution across participants. I used the same methods in this dissertation to generate random preference models of intertemporal choice except for the replacement of truncated normal distributions between 0 and 1 with beta distributions. The support of beta distributions is actually between 0 and 1 and thus they are a natural choice for random variables within this range.

Proportional difference (PD) model. The proportional difference model (González-Vallejo, 2002) provides an alternative way to model choice variability in empirical data. Specifically, this model assumes that, when facing a binary choice between options with multiple attributes, people first evaluate the proportional difference between options within each attribute and then rely on a linear combination of the proportional differences to make a choice. In other words, the PD model adopts an attribute-wise approach to choice behavior as opposed to the alternative-wise approach of utility-based models. When intertemporal choice is of concern, this model suggests that the proportional difference between a smaller amount, v_s, and a larger amount, v_l, equals $\frac{v_l - v_s}{v_l}$ and the proportional difference between a shorter delay, t_s, and a longer delay, t_l, equals $\frac{t_l - t_s}{t_l}$. The first difference can be interpreted as the relative advantage of an LL option, (v_l, t_l), over an SS option, (v_s, t_s) on the money attribute while the second can be interpreted as the relative advantage of the SS option over the LL option on the delay attribute. Therefore, we can subtract the second proportional difference from the first one to generate an overall evaluation of the LL option; the overall evaluation of the SS option is just the opposite. This way of evaluating options differs fundamentally from the alternative-wise method involved in the delay discounting paradigm. With an additional assumption of a normally distributed random disturbance to the overall evaluation. the PD model entails that

$$P(LL|\{SS, LL\}) = \Phi(\frac{d-\delta}{\sigma}) \tag{27}$$

in which $d = \frac{v_l - v_s}{v_l} - \frac{t_l - t_s}{t_s}$ represents the overall evaluation of the LL option without disturbance, δ is a parameter on personal decision threshold, σ is the standard deviation of the normally distributed random disturbance term, and Φ is the cumulative distribution function of a standard normal distribution as in Probit models (i.e., Equation 23). Both δ and σ are free parameters in the PD model and δ can also be interpreted as the relative significance of the delay attribute to the money attribute (González-Vallejo & Reid, 2006).

Although both the PD model and the Probit models of intertemporal choice discussed above involve a normally distributed disturbance or error term, they are different in two important aspects. First, the variable d in Equation 27 is calculated from within-attribute differences, i.e., the proportional differences in money and delay, whereas the same variable in Equation 23 is determined by within-alternative quantities, that is, the (discounted) utilities of the two options. Second, the PD model involves one more parameter, the personal decision threshold, which denotes the degree to which a decision maker differentially weighs the proportional differences on the two attributes. With this extra parameter, the PD model can assign unequal weights to the two attributes, whereas the Probit models always weigh the discounted utilities of both options equally. Consequently, the PD model is more flexible than the Probit models.

As mentioned above, the proportional differences in the PD model are similar to the relative differences in the attribute-wise models assuming an attention shift mechanism in the sense that they both evaluate within-attribute differences in a proportional way. Consequently, both of them can be interpreted as the relative advantages or disadvantages of one option over the other. Nonetheless, unlike the PD model, the attribute-wise models with relative differences in Dai and Busemeyer (2014) used local minima (i.e., v_s and t_s)

instead of local maxima (i.e., v_l and t_l) to normalize the direct differences. The rationale for this modification is that in an intertemporal choice scenario people tend to treat the SS option as the reference point (Weber et al., 2007) and thus the associated attribute values, i.e., v_s and t_s, are more likely to serve as the normalizers. Dai and Busemeyer examined both the attribute-wise models with relative differences and the original PD model. It turned out that both classes of models did not perform well in describing the empirical data. In this dissertation I test these models against different types of intertemporal choice questions to check whether the inferiority of models using relative differences is stimulus-dependent or not.

Tradeoff model. Like the PD model, the tradeoff model of intertemporal choice (Scholten & Read, 2010; Scholten, Read, & Sanborn, 2014) also assumes an attribute-wise approach. Specifically, it uses two intra-attribute weighing functions to measure the direct difference between valued outcomes and direct difference between weighted delays, and two inter-attribute weighing functions to make tradeoff between the direct differences. The two intra-attribute weighing functions are a value function, $u(v)$, similar to the value transformation function in the DFT models in Dai and Busemeyer (2014) and a time weighing function, $w(t)$, similar to the time transformation function in the DFT models. On the one hand, both models assume that within-attribute differences are taken in a direct way after performing certain forms of transformation on the objective money amounts and delay durations. On the other hand, the tradeoff model uses a logarithmic transformation instead of a power transformation as in the DFT models. After getting direct differences between valued outcomes and weighted delays, the tradeoff model then weighs the differences on money and delay attributes against each other with the inter-attribute weighing function $Q_{T|V}$ and $Q_{V|T}$. Specifically, according to a deterministic version of the tradeoff model, one should prefer the LL option if

$$Q_{T|V}(w(t_l) - w(t_s)) < Q_{V|T}(u(v_l) - u(v_s)) \tag{28}$$

and prefer the SS option if

$$Q_{T|V}(w(t_l) - w(t_s)) > Q_{V|T}(u(v_l) - u(v_s)). \tag{29}$$

Obviously, this deterministic model of intertemporal choice is similar to the attribute-wise core theory with direct differences mentioned above.

To extend the deterministic tradeoff model to accommodate noisy choice data, Scholten and Read (2010) proposed the following ratio rule on choice probability:

$$Pr(LL|\{SS, LL\}) = \frac{g}{g+f} \tag{30}$$

in which $g = Q_{V|T}(u(v_l) - u(v_s))$, the weighted advantage of the LL option on the money attribute, and $f = Q_{T|V}(w(t_l) - w(t_s))$, the weighted advantage of the SS option on the time attribute. To implement Equation 30 in practice, we still need to specify the form of $Q_{T|V}$ and $Q_{V|T}$. The latest development of the tradeoff model (Scholten et al., 2014) uses the following formula to specify the odds of choosing the LL option:

$$\Omega_{LL} = \left(\frac{\frac{1}{\gamma}(log(1+\gamma v_l) - log(1+\gamma v_s))}{\frac{\kappa}{\alpha}log(1+\alpha(\frac{\frac{1}{\tau}(log(1+\tau t_L)-log(1+\tau t_s))}{\ell})^{\ell})}\right)^{1/\epsilon}. \tag{31}$$

In Equation 31, $\gamma, \tau > 0$ are the diminishing marginal sensitivity parameters for money and delay; $\kappa > 0$ is a tradeoff parameter serving the same role as the concept of discounting or impatience; $\alpha > 0$ is a parameter of subadditivity; $\ell > 1$ is a parameter for superadditivity; and ϵ is a noise parameter. With this set of parameters, the probabilistic tradeoff model can explain a large number of effects and phenomena in intertemporal choice, including the magnitude effect, the common difference effect, subadditivity and superadditivity in delay discounting, and intransitivity. However, this model is silent on the underlying processes leading to the explicit choice and thus says nothing on response time distribution. One major purpose of this dissertation is to develop a dynamic model that can not only make the same qualitative predictions as the tradeoff model when considering choice probabilities but also account for the response times associated with different choices.

Diffusion models. All the models introduced above are static because they do not provide an explicit account of the underlying processes culminating in explicit choices. Therefore, they are capable of predicting the choice probability of each option but not the associated decision times. This disadvantage of static models makes them less desirable than dynamic models since the latter can explain not only the choice probabilities but also the decision times for each choice question. Consequently, Dai and Busemeyer (2014) developed and tested a large group of dynamic models of intertemporal choice built upon diffusion processes. Such dynamic models are specific cases of sequential sampling models which assume an evidence or preference accumulation process with continuous time, varying amounts of evidence, and a relative stopping rule. Specifically, this class of models suggest that evidence for or against each option is sampled and accumulated sequentially over time as people deliberate on the choice question. This accumulation process continues until the evidence or preference level of one option reaches a threshold first. At that time, a decision is made to choose the very option whose accumulated preference level has reached the threshold. Furthermore, the deliberation or decision time and the time required for non-decisional components in the choice task combined determine the actual response time. Because the preference accumulation process is a stochastic process with varying amounts of instantaneous change in preference over a continuous time scale, there are an infinite number of possible accumulation paths, making the resultant choice and decision time inherently probabilistic. Therefore, such models usually make probabilistic rather than deterministic predictions on choices and decision times (and thus response times). See Ratcliff and Smith (2004) for a comparison of diffusion model and other types of sequential sampling models for binary choice tasks.

Usually there are five parameters involved in a diffusion model for a binary choice task such as that used in most traditional studies on intertemporal choice. The first parameter is the mean drift rate (or the drift parameter), d, which indicates the average rate of evidence accumulation for a certain option. The higher it is, the more quickly evidence for the option accumulates. The second parameter is the diffusion parameter, σ, which reflects the degree of variability in instantaneous rate of evidence accumulation. A non-zero value of σ is usually required for a probabilistic model of choice behavior because otherwise the trajectory of preference accumulation will be a straight line solely determined by the mean drift rate. This will lead to deterministic choice and decision time given that all parameters are assumed to be fixed across trials. The third parameter, θ, denotes the threshold level on preference strength. The higher it is, the longer it takes to reach the threshold. As a result, it provides a mechanism for explaining speed-accuracy trade-off in binary choice tasks. Specifically, with a higher threshold, people become more likely to choose the option preferred by the deterministic core theory, but it takes more time to reach this decision. To the contrary, when the threshold is relatively low, people tend to make quick choices but the probability of choosing the inferior option according to the core theory increases.

The fourth parameter, z, represents the initial preference level before the preference accumulation process gets started. It can be interpreted as a measure of response bias toward a specific option. Specifically, an option with a positive initial preference level is more likely to reach the preference threshold first than the other option when all the other factors are equal. The last parameter, T_{er}, represents the amount of non-decisional time involved in the task, and it is required for predicting response time distribution. Diffusion models for binary choice tasks using these five parameters actually assume the Wiener process as the underlying stochastic process leading to explicit choices. Given the values on d, σ, θ, and z for a specific option, its binary choice probability equals (Busemeyer & Diederich, 2009; Ratcliff, 1978)

$$Pr(d, \sigma, \theta, z) = \frac{1 - exp(-2d(\theta + z)/\sigma^2)}{1 - exp(-4d\theta/\sigma^2)}. \tag{32}$$

Furthermore, the probability density of a certain response time, t , given a chosen option with the above parameter values equals (Busemeyer & Diederich, 2009; Ratcliff, 1978)

$$P(t) = \pi \cdot (\frac{2\theta}{\sigma})^{-2} exp(\frac{d(\theta - z)}{\sigma^2}) \cdot \sum_{j=1}^{\infty} j \cdot exp(-\frac{1}{2}((\frac{\pi j \sigma}{2\theta})^2 + (\frac{d}{\sigma})^2) \cdot (t - T_{er})) \cdot sin(\frac{\pi(\theta - z)j}{2\theta}). \quad (33)$$

The summation in Equation 33 involves an infinite number of terms. In practice it has to be evaluated approximately by choosing a lower threshold on the term

$$\pi \cdot (\frac{2\theta}{\sigma})^{-2} \cdot j \cdot exp(\frac{d(\theta - z)}{\sigma^2} - \frac{1}{2}((\frac{\pi j \sigma}{2\theta})^2 + (\frac{d}{\sigma})^2) \cdot (t - T_{er})). \quad (34)$$

In Dai and Busemeyer (2014) and this dissertation, the lower threshold was set at 10^{-200} for the approximation.

When addressing the issue of intertemporal choice, a large number of diffusion models with different stochastic specifications can be developed by setting the relevant parameters in distinct ways. First of all, like the Probit models discussed above, we can set parameter σ to be independent or dependent on attribute values. In the independent case, σ is a free parameter to be estimated from data as in the first class of Probit models. By contrast, when σ is set to be associated with attribute values, Equations 24 - 26 can be used to calculate its value. In this way, the value of σ varies from question to question. Consequently, I hereafter call the first class of Probit models and diffusion models (i.e., with independent σ) as models with fixed σ, and the second class of Probit and diffusion models (i.e., with σ dependent on attribute values) as models with varied σ.

Second, we can either set the response bias parameter, z, to zero to generate simple unbiased diffusion models or let it be a free parameter fixed across trials to create more complex models. The former is a special case of the latter and thus more parsimonious. Furthermore, we can also try a special class of diffusion models in which parameter z has a fixed absolute value across trials but its sign changes from trial and trial contingent on the sign of parameter d. Specifically, the two parameters always have the same sign so that the direction of the initial bias is always in favor of the option with a positive d value. This specific type of diffusion model predicts that, for each pair of options, the one with a higher choice probability (i.e., above .5) also tends to have shorter response times. Of course, as a specific case of diffusion models, such models also predict that extreme choice probabilities are associated with shorter response times across different pairs of options. This combination of response time patterns might not be common in empirical data, but when revealed, it imposes a strict constraint for choosing the appropriate form of diffusion models. In summary, there are three different ways to set parameter z to generate different diffusion models, i.e., always at zero, as a free parameter fixed across trials, or as a free parameter with a fixed absolute value but varying signs across trials. Hereafter I call the first class of diffusion models as diffusion models without initial bias, the second diffusion models with fixed initial bias or z, and the third diffusion models with varied initial bias or z.

Finally, we can set θ either proportional to σ or as a fixed parameter to be estimated from data. Dai and Busemeyer (2014) chose the first approach and thus treated $\theta^* = \theta/\sigma$ as the free parameter in their model fitting procedure. The rationale for this approach is that people tend to be more cautious when there is high variability in instantaneous evaluation, i.e., a large σ. In the context of diffusion models, this implies a high threshold on preference level, i.e., a large θ. In other words, there seems to be a proportional relationship between θ and σ. Because the absolute value of parameter z should not exceed the value of parameter θ and it is a common practice to set z proportional to θ, when θ is proportional to σ, the same should also be applied to parameter z . In this case, $z^* = z/\sigma$ was actually the free parameter to estimate.

However, when θ is proportional to σ and z equals 0, the corresponding diffusion models built upon the attribute-wise core theory with direct differences entail weak stochastic transitivity (WST). Because the winning model in Dai and Busemeyer (2014) was actually such a model, it also predicted WST. If intertemporal choice is indeed intransitive in the sense of WST, it is necessary to modify the winning model in Dai and

Busemeyer to incorporate this property. One possible method is to relax the proportional assumption so that θ and z are treated as constants across pairs of options and thus free parameters in the model. In this way, the predicted choice probability is an increasing function of d/σ^2 instead of d/σ when θ is proportional to σ, and this makes it possible to predict a violation of WST under the general framework of DFT. Therefore, in this dissertation, I also explored diffusion models with fixed θ and z to examine their capability of describing empirical data and accounting for potential intransitivity in intertemporal choice.

Yet another way to make the winning model in Dai and Busemeyer (2014) capable of predicting a violation of WST is to modify its core theory. When the building block of the core theory is subjective perceptions of the (objective) direct differences in value and time and the transformation functions connecting objective and subjective differences is non-linear, the resultant DFT models do not necessarily predict WST. Here by subjective perceptions of the direct differences I mean $u(v_l - v_s)$ and $p(t_l - t_s)$. In this way, the model takes direct differences in objective amounts before applying transformations. In other words, the model uses subjective perceptions of objective direct differences to calculate d and σ rather than takes direct differences in subjective value and time as the building blocks of d and σ. Due to the different way of calculating d and σ, the resultant diffusion model can predict a violation of WST even if $\theta^* = \theta/\sigma$ and $z^* = z/\sigma$ are treated as free parameters.

However, simply replacing direct differences in subjective value and time with subjective perceptions of direct differences in (objective) value and time leads to a qualitative drawback of the new model. Specifically, such a model cannot predict the common difference effect because $p(t_l - t_s)$ will not change under the manipulation for the common difference effect. However, the choice proportions did change in empirical data under the manipulation, at least for a portion of participants. A remedy for this problem is to include both direct differences in subjective value and time and subjective perceptions of direct differences in (objective) value and time in the core theory so that the resultant DFT model can predict both common difference effect and a potential violation of WST. Specifically, the corresponding core theory suggests that one should choose the LL option if

$$d = w \cdot [k \cdot (u(v_l) - u(v_s)) + (1-k) \cdot u(v_l - v_s)] - (1-w) \cdot [k \cdot (p(t_l) - p(t_s)) + (1-k) \cdot p(t_l - t_s)] \quad (35)$$

is positive, and choose the SS option if d is negative. The k parameter in Equation 35 ranges from 0 and 1 and it determines the relative weights of the two types of direct differences between options. If $k = 0$, the model suggests that people take into account only subjective perceptions of direct differences in objective money and time. To the contrary, when $k = 1$, the model implies that people consider only direct differences in subjective value and time. I will thereafter call models of this type attribute-wise models with mixed direct differences. Of course the validity of such models still needs to be tested against empirical data.

To sum up, I use two different treatments of σ (fixed or varied), three different treatments of z (fixed at zero, freely estimated, or freely estimated but having the same sign as parameter d), and two different treatments of θ(proportional to σ or not, also applied to z when it is not fixed at zero) to factorially generate 12 stochastic specifications for diffusion models. As in other classes of probabilistic models of intertemporal choice, the parameter d is calculated from the core theories (i.e., Equations 15-18). I also examine the attribute-wise models with mixed direct differences coupled with two different stochastic specifications based on diffusion processes to check whether the resultant models perform even better than the best model with only direct differences in subjective value and time. Specifically, I explore the stochastic specification used in the previous winning model (i.e., a DFT model with varied σ, fixed θ^* and no initial bias) and that with varied σ, fixed θ^*, and varied z^* with the same sign as the d parameter.

Summary of All Models Examined in This Dissertation

In total, I examine 379 probabilistic models of intertemporal choice in this dissertation. Specifically, I factorially combine the six pairs of value and time transformation functions (i.e., Equations 9-14), the four core theories (i.e., Equations 15-18), and the 17 stochastic specifications (i.e., those of constant error models, Probit models with fixed σ, Probit models with varied σ, logistic models, random preference models, and the 12 stochastic specifications based on diffusion process) to generate 374 models. Note that for models with the core theory using relative differences, including a scaling constant in the time transformation function does not make any difference (i.e., Equation 10 was equivalent to Equation 11 in this case and the same was true between Equations 12 and 13). Consequently, there are only 68 models built upon relative differences instead of 102 models based on direct difference, exponential discount function, or hyperbolic discount function. In addition, I also explore two probabilistic versions of the tradeoff models (with the noise parameter ϵ fixed at 1 or freely estimated), the PD model, the diffusion model with mixed direct differences, varied σ, fixed θ^* and no initial bias, as well as the diffusion model with mixed direct differences, varied σ, fixed θ^*, and varied z^* having the same sign as d. To distinguish between previous diffusion models based on Equation 17 and those with mixed direct differences, I call the former diffusion models with single direct differences. To reveal the results a little earlier, the last model in general performs the best with regards to various model selection methods.

Summary of Goals for This Dissertation

The pivotal goal of this dissertation is to extend the work of Dai and Busemeyer (2014) to find a model that can not only qualitatively account for a wide range of effects and phenomena in intertemporal choice but also quantitatively fit and predict empirical data better than other models. To this end, I first used the data from Experiment 3 in Dai and Busemeyer to examine the extended repertoire of probabilistic models of intertemporal choice. I chose this set of data because the relevant experiment has the most refined design among the three experiments reported and it provided the strongest support for the probabilistic nature of intertemporal choice. After that, I conducted two new experiments with intertemporal choice questions differing from those in Dai and Busemeyer to generate more data for testing the models. Specifically, intertemporal choice questions with immediate SS options were included in the first experiment and question sets for testing intransitivity of intertemporal choice were used in the second experiment. The purpose of the first experiment was to test whether attribute-wise models also performed better than alternative-wise ones even when SS options occurred immediately. The purpose of the second experiment was to examine transitivity of intertemporal preference at an individual level so that more informative data could be collected for comparing the tradeoff models with DFT models, among others. These new experiments and the previous one combined covered most of the effects and phenomena in intertemporal choice. Consequently, one could have more confidence in the winning model across all three studies, if any.

General Method

Stimuli

As most studies on intertemporal choice, experiments reported in this dissertation used binary choice questions to examine people's intertemporal preference. Two measures were taken in Experiment 3 of Dai and Busemeyer (2014) and the two new experiments to produce data suitable for testing and comparing different probabilistic models,. First, because substantial individual differences in discount rate had been revealed in the literature, different set of intertemporal choice questions was generated for each participant to avoid extreme choice patterns, i.e., choosing the SS (or LL) options all the time. Specifically, for each participant in a study, an adjustment procedure was used at an early stage of the study to generate one or more pairs of approximately indifferent options, on the basis of which formal choice questions were created. Because the approximately indifferent pair(s) tended to differ among participants, the resultant formal questions also differed among participants. Here by approximately indifferent I mean that the two options in the pair had about the same choice probability instead of suggesting that intertemporal preference is essentially deterministic.

Second, each question was presented repeatedly in these experiments to generate data particularly useful for examining probabilistic models and testing weak stochastic transitivity in intertemporal choice at an individual level. To my best knowledge, each question was presented to a participant only once in every previous study on intransitivity of intertemporal preference and the related phenomenon of nonadditivity in delay discounting. Consequently, when choice variability was taken into account, only an aggregate-level analysis was possible and the resultant conclusion might be misleading when applied to individual participants. This was clearly not satisfying, and thus in this dissertation I adopted the new design to run a more rigorous test on the issue of intransitivity in intertemporal choice.

Participants

Participants were volunteers recruited at Indiana University via advertisement on notice boards across campus. They were paid for their participation and the amount and date of payment for each participant was usually determined by his/her choice in one question randomly picked from his/her question set. Specifically, each participant would get one-fourth of the amount he/she chose in the randomly picked question in addition to a base payment of 4 dollars, and he/she needed to wait the period of time associated with the option he/she chose. For example, if a participant chose receiving 20 dollars in a week as opposed to receiving 10 dollars now in the randomly picked question, then he/she would get 9 (.25*20 + 4) dollars one week after taking the study.

Procedure

Each experiment in this dissertation included a practice session, an adjustment procedure and a subsequent formal session. The computer-based experimental setting was fulfilled by a set of MATLAB programs, which used MATLAB base functions and the Cogent toolbox to record both choice responses and response times. At the beginning of each experiment was the practice session through which each participant could get familiar with the choice task. After that, an adjustment procedure was implemented to find approximately indifferent pair(s) for each participant, on the basis of which formal questions were generated. The adjustment procedure used either single lower and upper limits [in Experiment 3 of Dai and Busemeyer (2014)] or double lower and upper limits (in the new experiments). See Dai and Busemeyer (2014) for details of the single limit method and see Richards, Zhang, Mitchell, and DeWit (1997) for details of the double-limit method. Finally, in the formal session of each experiment, questions were presented repeatedly and in a random order to collect choice and response time data. Major and contingent short breaks were included in each experiment in order to reduce the fatigue effect. Furthermore, equally-spaced filler questions were interspersed among formal questions to detect inattention. Specifically, each filler question contained a pair of dominating and dominated options, i.e., one with a larger reward but a shorter delay and the other one with a smaller reward but a longer delay. If a participant chose the dominated option in a filler question, a warning sign would pop up asking for more attention and an optional short break would be given. Options were presented on either the left-hand side or the right-hand side of the screen and participants indicated their choices by clicking the mouse button on the same side. The positions of SS and LL options were randomized across trials to prevent participants from habitually clicking the same mouse button all the time. Participants were instructed to consider each piece of information carefully for each question. The payment plan was emphasized in both the consent form, which participants signed before starting the experiments, and the instructions involved in the experiments.

See Figure 1 for a screenshot of the experimental setting in the three experiments.

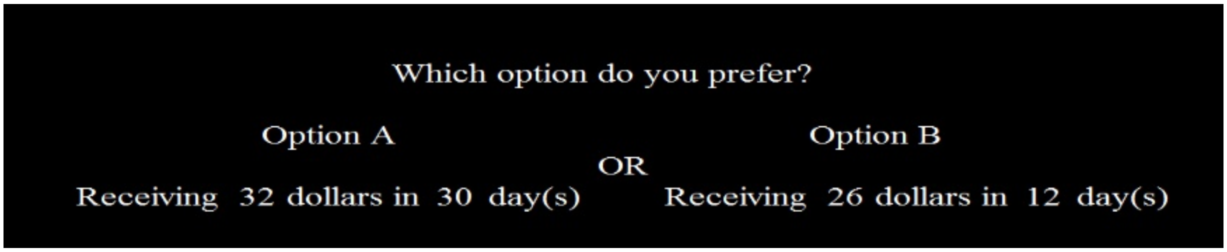

Figure 1. Screenshot of the experimental setting. The positions of SS and LL options were randomized across trials.

Data Analysis

To analyze the data and fit various probabilistic models to them, I first pruned the data to remove choices with extreme response times. Specifically, any choice with a response time shorter than 1500ms or longer than 10000 ms was excluded from further analyses. The lower bound was set to ensure that participants spent sufficient amount of time sampling all the information in each question, and the upper bound was chosen to avoid the disproportional impact of extremely long response times on parameter estimation when response time was taken into account. The extreme long response times also suggested that something unusual happened in the decision process and thus the associated choices might distort the results if included in further analysis. It turned out that most response times were within the the range set by the lower and upper bounds.

Choice variability as an indication of probabilistic nature of intertemporal choice. To corroborate the conclusion concerning probabilistic nature of intertemporal choice in Dai and Busemeyer (2014), I first examined choice variability within each repeatedly presented question from each experiment. If one switched between the SS and LL options when the same intertemporal choice question was asked multiple times, then we could properly conclude that intertemporal choice was probabilistic for the very person. Specifically, for each participant, I calculated the choice proportion of the LL option across repeated presentations of each question and counted the number of questions with moderate choice proportions (i.e., between 0 and 1 exclusively) as an indication of choice variability and probabilistic nature of intertemporal choice at an individual level. The result of such analysis for Experiment 3 in Dai and Busemeyer was already reported and those from the new experiments provided more support for the general conclusion.

Relationships between choice proportions and response times. Dynamic nature of intertemporal choice is another property I want to substantiate in this dissertation. This property suggests that intertemporal choice is governed by a decision process that takes different amounts of deliberation time for different choice questions. One convenient way to gather evidence for this property is to examine the relationships between choice proportions and response times as in Dai and Busemeyer (2014). This is due to the fact that a static perspective on intertemporal choice makes no non-trivial prediction on such relationships, while different formulations of the dynamic process leads to different relationships. Consequently, detection of any non-trivial relationship between choice proportions and response times in observed data would not only support the dynamic perspective but also suggest specific types of decision process. For example, dynamic diffusion models generally predict that extreme choice probabilities tend to be coupled with shorter response times and vice versa. Empirically, this suggests an inverse U-shaped relationship between choice proportions of the LL options and the mean response times across different choice questions. Because a large number of diffusion models were to be tested in this dissertation, it was necessary to first examine this general prediction of diffusion models. If this prediction did not hold in the empirical data, then it was untenable to assume diffusion process as the underlying mechanism of intertemporal choice.

To show that the relationship suggested by diffusion models did exist in observed data, for each participant I calculated the choice proportion of the LL option and the mean response time for each intertemporal choice question presented repeatedly. The results of all participants and questions were then categorized into five equal-interval bins with regard to actual choice proportion, and the mean response time averaged across all participants and questions for each bin of choice proportion was then graphed to show the inverse U-shaped relationship. In addition, for each participant I averaged the mean response times for questions with extreme choice proportions (i.e., below 0.2 or above 0.8) and those for questions with moderate choice proportions (i.e., between 0.2 and 0.8) to get two related measures of response time. With the related measures on response time associated with different choice proportions, a related-samples t test was performed to check whether the inverse U-shaped relationship held in the empirical data. I also compared the average mean response times for questions with extreme and moderate choice proportions for each individual to show that the inverse U-shaped relationship was not merely an aggregate artifact. Hereafter I will call this type of relationship between choice proportions and response times a marginal relationship since the response times for both options in each question were averaged to get a marginal measure. Such an analysis was involved in Dai and Busemeyer (2014) and the relevant result for Experiment 3 has been reported. I reported the result again in this dissertation in addition to those from the new experiments to make the conclusion more convincing since the previous study examined different aspects of intertemporal choice (i.e., the delay duration effect, the common difference effect, and the magnitude effect) that were not covered in the new experiments.

There might be another relationship between choice proportions and response times in intertemporal choice. That is, for each question presented repeatedly, the option chosen more frequently is also associated with a shorter response time on average than the other option. Because conditional response times were averaged to get a measure for each option in this analysis, I will call it a conditional relationship from now on. This potential relationship between choice proportions and response times was not investigated in Dai and Busemeyer (2014) but it suggests a specific form of diffusion models, i.e., the diffusion models with bias parameter, z, having the same sign as the mean drift rate parameter, d. If such a relationship actually arises in empirical data, then we have a reason to try the specific type of diffusion models and they are likely to win the model comparison discussed below. To check whether this relationship existed in the actual data, for each individual I first categorized either option in each question as the more frequently chosen or less frequently chosen option and calculated corresponding mean response times. This was only possible when a participant switched between the SS and LL options within the question and the two options did not have the same choice proportion (i.e., 0.5). Consequently, those questions whose options could not been properly categorized were excluded from further analysis. The next step in the analysis was to calculate the average mean response times associated with more frequently chosen options and less frequently chosen options for each participant to get two related measures. As for the marginal relationship, I then ran a related-samples t test to examine the conditional relationship between choice proportions and response times. I also reported a graph for the aggregate result and conducted individual-level analyses to show that the revealed relationship was not merely an aggregate artifact.

Preliminary model fitting and comparisons. The principal model fitting method used in this dissertation was the maximum-likelihood estimation (MLE). Specifically, for each participant, the log likelihood of the actual result on each trial predicted by a model using the exact money amounts and delay durations was summed across trials to produce the summed log likelihood denoted as LogL. After that, the SIMPLEX algorithm implemented by the fminsearch function in MATLAB was employed to find the maximum-likelihood estimates of the relevant parameters for each participant. Because both static and dynamic models were included in the repertoire but the former can predict only choice probabilities, I first fit all the models to individual choice data as the initial step toward comparing static models against dynamic ones and comparing alternative-wise models against attribute-wise ones.

Because the various models examined in this dissertation differed in complexity with regard to number of parameters and other aspects such as functional forms, it was necessary to take model complexity into account when comparing these models. One convenient and widely adopted way to address the issue of model complexity is to use model comparison indices based on MLE such as the Akaike Information Criterion (AIC; Akaike, 1974) and the Bayesian Information Criterion (BIC; Schwarz, 1978) . Each of the indices applies a punishment term based on number of parameters to the maximum log-likelihood of data given a model to compensate for the differential flexibility of various models. Specifically,

$$AIC = -2Log(\hat{L}) + 2k, \tag{36}$$

and

$$BIC = -2Log(\hat{L}) + log(n) \cdot k. \tag{37}$$

In Equations 36 and 37, $Log(\hat{L})$ represents the maximum log-likelihood of data given a model, k is the number of free parameters involved in the model, and n is the number of data points involved in the calculation of $Log(\hat{L})$. For the data analyzed in this dissertation, n equaled the number of trials involved in the calculation of $Log(\hat{L})$ because each trial contributed a data point whose log-likelihood was summed. In general, a lower value on either AIC or BIC suggests a better model that balances between goodness-of-fit and model complexity. For each selection criterion, there were three possible ways to choose the best model, i.e., the overall criterion value for a model, the count of lowest criterion values across participants for a model, and the result of pairwise comparisons (Broomell, Budescu, & Por, 2011). Specifically, the first method indicated the general performance of a model when fitting all individual data, the lower the better. The second showed the number of participants whose data a specific model fit best, the higher the better. And the last method indicated the relative performance of a model when compared to another one in terms of the count of lower criterion values across participants. With the last method, it was possible to find a best model that produced a lower criterion value on more participants than any other model when only two models were compared each time.

In the preliminary analysis, I chose the first method to compare different models. The major reason for using this method was that it provided straightforward order among all the competing models, whereas count of lowest AIC or BIC values across participants might produce ties among models and pairwise comparisons could lead to circular relationships among models.

Note that, because AIC and BIC consider only one aspect of model complexity, i.e., number of parameters, they only provide a partial solution to the issue of model complexity when comparing models. Furthermore, because AIC and BIC differ in the punishment term and for relatively large data sets, $log(n)$ would be larger than 2, BIC tends to apply a heavier punishment than AIC. As a result, AIC and BIC may lead to different conclusions on the best model among a set of models. Specifically, AIC usually favors more complicated models while BIC tends to pick simpler ones. Therefore, I used overall AIC and BIC values as only the preliminary devices for comparing different models with regard to choice responses. After finding a small set of candidate models using these two criteria, I proceeded to compare the limited set of models in more sophisticated ways.

Reanalysis of Previous Data

In this section I report the results of reanalyzing the data from Experiment 3 in Dai and Busemeyer (2014). Some of the results were reported before but not the others. Therefore I only provide a summary of those already reported and elaborate on those new results. The major reason for including this section in the dissertation is to provide a comprehensive summary of evidence for the probabilistic and dynamic nature of intertemporal choice. On the one hand, Experiment 3 in Dai and Busemeyer was designed around the same effects in intertemporal choice as the other two studies but it had the most refined design. Therefore, it could be viewed as a representative study in the literature along the line of research on probabilistic and dynamic nature of intertemporal choice. On the other hand, Experiment 3 in Dai and Busemeyer was similar to the new experiments conducted for this dissertation in the sense that each of them involved repeatedly presented questions. Therefore, the same analysis framework could be employed across the three studies reported in this dissertation. More importantly, the three studies reported here covered most effects and phenomena in intertemporal choice between gains and thus provided a solid foundation on which a comprehensive model can be developed and tested. Note that the resultant model could be easily extended to intertemporal choice involving negative outcomes.

Method

Choice questions in Experiment 3 of Dai and Busemeyer (2014) were designed to examine the delay duration effect, common difference effect, and magnitude effect in intertemporal choice. In this study, each question was presented five times to reveal the probabilistic nature of intertemporal choice. Thirty-seven participants (26 females and 11 males) with an average age of 21 were recruited for the experiment. For each participant, a single-limit adjustment procedure was used to find three approximately indifferent pairs of options on the basis of which formal questions were generated. After that, each participant answered 300 formal questions, 100 for each effect. On average, participants received 12 dollars or so in about 35 days. See Table 1 for a sample of formal questions a typical participant answered in the experiment and Dai and Busemeyer for more information on this and the other two studies.

Results and Discussion

Choice variability as an indication of probabilistic nature of intertemporal choice. As reported in Dai and Busemeyer (2014), all participants switched between the SS and LL options in at least two unique questions associated with a specific intertemporal effect when the questions were presented repeatedly. Furthermore, 32 out of the 37 participants switched in at least two unique questions associated with the delay duration effect when the questions were presented repeatedly. These results suggested that intertemporal choice is probabilistic just like risky choice.

Table 1

Sample Formal Questions for a Typical Participant in Experiment 3 of Dai and Busemeyer (2014)

Question	Smaller reward (dollars)	Shorter delay (days)	Larger reward (dollars)	Longer delay (days)
		Delay duration effect		
1	20	2	36	4
2	20	4	36	8
3	20	6	36	12
4	20	8	36	16
5	20	10	36	20
		Common difference effect		
1	14	2	32	32
2	14	4	32	34
3	14	6	32	36
4	14	8	32	38
5	14	10	32	40
		Magnitude effect		
1	2	12	4	28
2	4	12	8	28
3	6	12	12	28
4	8	12	16	28
5	10	12	20	28

Relationships between choice proportions and response times. Figure 2 demonstrates the actual and predicted marginal relationships between choice proportions and response times in Experiment 3 of Dai and Busemeyer (2014). Specifically, the actual or predicted choice proportions of the LL options for repeatedly presented questions are categorized into five equal-size bins, and the corresponding actual or predicted average mean response times are shown on the vertical axis. Clearly there was an inverse U-shaped relationship between the actual choice proportions and actual mean response times across questions and participants. The difference in mean response time between questions with extreme and moderate choice proportions was statistically significant ($M_{extreme} = 3.89s$, $M_{moderate} = 4.30s$, $t[36] = -6.40$, $p < .01$). The same pattern arose in the individual data from 34 out of the 37 participants, with 10 reaching statistical significance. In summary, diffusion models' general prediction on the relationship between choice proportions and response times held in this experiment. I will discuss the predicted marginal relationships between choice proportions and response times at a later section after a best model across all studies is selected.

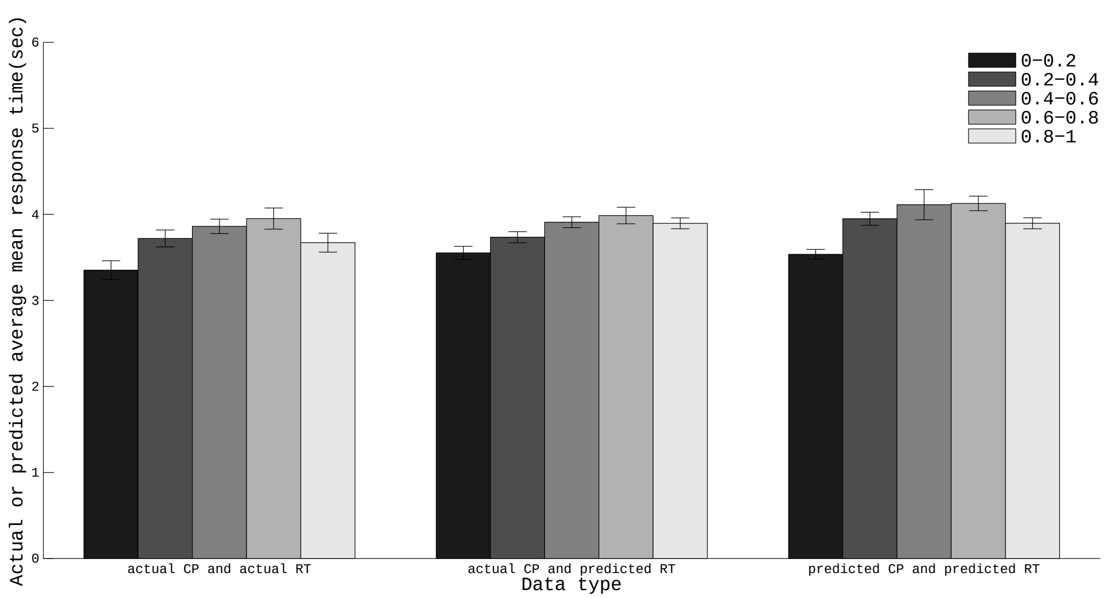

Figure 2. Marginal relationships between choice proportions and response times in Experiment 3 of Dai and Busemeyer (2014). Each bar is associated with a specific range of choice proportions of the LL options. Error bars show 95% confidence intervals.

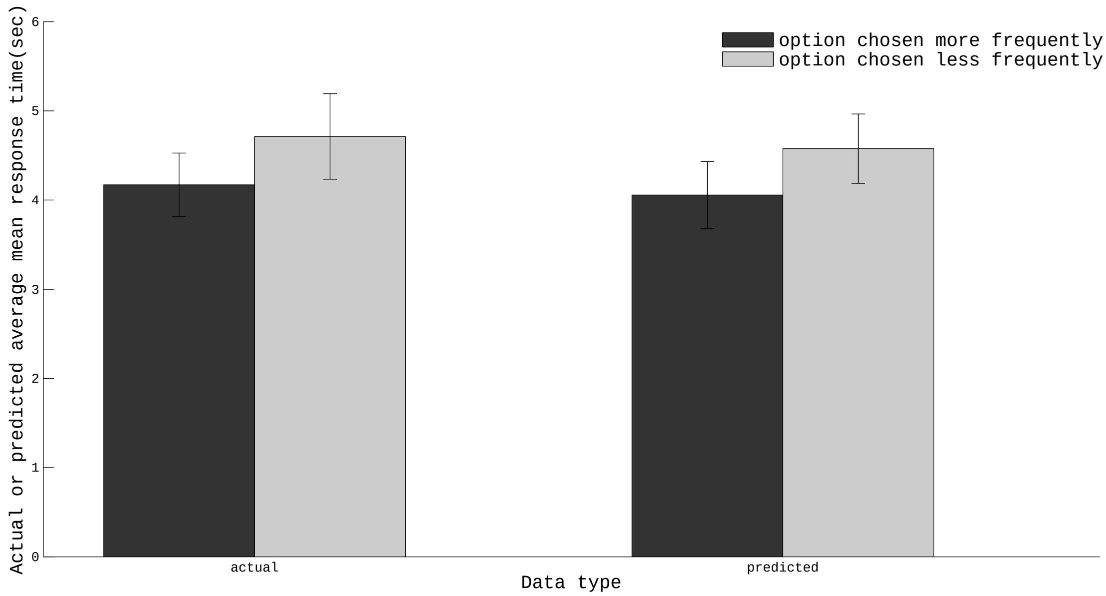

Figure 3. Conditional relationships between choice proportions and response times in Experiment 3 of Dai and Busemeyer (2014). Error bars show 95% confidence intervals.

Figure 3 shows the actual and predicted conditional relationships between choice proportions and response times. It is readily seen that response times for more frequently chosen options tended to be shorter than those for less frequently chosen options ($M_{more} = 4.17s$, $M_{less} = 4.71s$, $t[36] = -3.49$, $p < .01$). The same pattern was also found in individual data from 26 out of the 37 participants, with 7 reaching statistically significance. These results suggested that a specific form of diffusion models, i.e., those with an initial bias toward the option with a positive mean drift rate, may well account for the response time data. Again I will discuss the predicted conditional relationship between choice proportions and response times after selecting a best model across all studies.

Preliminary model fitting and comparisons. For a preliminary evaluation of model performance, I first fit all the probabilistic models to individual choice data and employed both AIC and BIC to compare the models. Table 2 lists the three best models in terms of overall AIC and BIC values. Clearly, all of the best models were built upon an attribute-wise core theory with direct differences and the best model in terms of overall BIC value matched the one reported in Dai and Busemeyer (2014) when a smaller set of models were fit and compared. Additionally, most of the best models assumed a diffusion process and thus had a dynamic structure. Because AIC and BIC involve different punishment terms, it is not unexpected that they suggested different best models. I will leave the decision on the single best model to a later section in which results from all studies are combined to get an overall evaluation of model performance.

Table 2

The Three Best Models for Choice Data in Experiment 3 of Dai and Busemeyer (2014) in Terms of Overall AIC Value or Overall BIC Value across Participants

Number	Transformations of objective value and time	Core theory	Stochastic specification	Number of free parameters	Overall criterion value
			AIC		
1	Logarithm	Attribute-wise with single direct differences	Diffusion model with varied σ, fixed θ^* and varied z^*	5	5750
2	Power with exponents no larger than 1	Attribute-wise with single direct differences	Diffusion model with varied σ, fixed θ^* and varied z^*	5	5792
3	Power with exponents no larger than 2	Attribute-wise with mixed direct differences	Diffusion model with varied σ, fixed θ^* and no initial bias	5	5795
			BIC		
1	Power with exponents no larger than 2	Attribute-wise with single direct differences	Diffusion model with varied σ, fixed θ^* and no initial bias	4	6982
2	Power with exponents no larger than 2	Attribute-wise with single direct differences	Random preference model	4	6985
3	Logarithm	Attribute-wise with single direct differences	Diffusion model with varied σ, fixed θ^* and no initial bias	4	7058

Experiment 1

Method

Sixty-nine participants (47 females and 22 males) with an average age of 20 were recruited for this experiment. On average participants received about 12 dollars in 28 days. Data were missing for one participant due to some technical problem and seven participants produced extreme choice responses, i.e., always choosing the LL options or the SS options across the whole stimulus set. Such type of data provided little information for revealing properties of intertemporal choice and distinguishing among competing models. Therefore, they were excluded from further analysis. Among the remaining 61 participants, there were 42 females and 19 males with the same average age as the original sample. The average payment was virtually the same as the original sample and the average payment delay was about 22 days.

In this experiment, at least half of the formal questions each participant answered involved an immediate SS option. Specifically, 30 out of the 61 participants who contributed valid data were required to choose between immediate SS options and delayed LL options throughout the study, while the rest needed to choose between immediate SS options and delayed LL options on half of the trials and between delayed SS and LL options on the other half. Each question with a delayed SS option was generated by increasing the delay durations of both options in a question with an immediate SS option to the same degree. Therefore, the resultant pairs of questions could be utilized to examine the common difference effects as well. For each participant, a double-limit adjustment procedure was implemented to find a pair of approximately indifferent options, on the basis of which formal questions were generated. For each participant, there were 48 unique formal questions and each was presented five times. Compared with the stimuli involved in Dai and Busemeyer (2014), the current type of stimuli was more similar to those used in traditional studies on intertemporal choice whose results favored the hyperbolic discounting model. Consequently, I ran this experiment to examine whether the same conclusion still held when the data were analyzed from a probabilistic perspective. If it turned out that in general attribute-wise models performed better than alternative-wise ones in fitting the resultant data, then we could have more confidence in the winning attribute-wise model. Furthermore, for each participant who encountered both immediate and delayed SS options, a Wilcoxon rank-sum test was performed on the choice proportions of LL options to examine whether the common difference effect appeared or not. See Table 3 for a sample of formal questions a typical participant encountered in this experiment.

Table 3

Sample Formal Questions for a Typical Participants in Experiment 1

Question	Smaller reward (dollars)	Shorter delay (days)	Larger reward (dollars)	Longer delay (days)
Always immediate SS options				
1	25	0	33	12
2	25	0	33	13
3	25	0	33	14
4	25	0	33	16
5	25	0	33	17
6	25	0	33	18
Mixture of immediate and delayed SS options				
1	25	0	33	12
2	25	0	33	13
3	25	0	33	14
4	25	10	33	22
5	25	10	33	23
6	25	10	33	24

Results

Choice variability as an indication of probabilistic nature of intertemporal choice. As before, each participant in the current study switched between the SS and LL options in one or more questions when they were presented multiple times. On average, participants changed their mind in about 16 out of 48 unique questions presented repeatedly.

Relationships between choice proportions and response times. Figure 4 demonstrates the actual and predicted marginal relationships between choice proportions and response times in the current study. As before, an inverse U-shaped relationship between actual choice proportions and mean response times was revealed across questions and participants ($M_{extreme} = 3.57s$, $M_{moderate} = 4.01s$, $t[58] = -8.84$, $p < .01$). The same pattern was found in individual data from 55 out of the 61 participants, with 37 differences reaching statistical significance. [4]

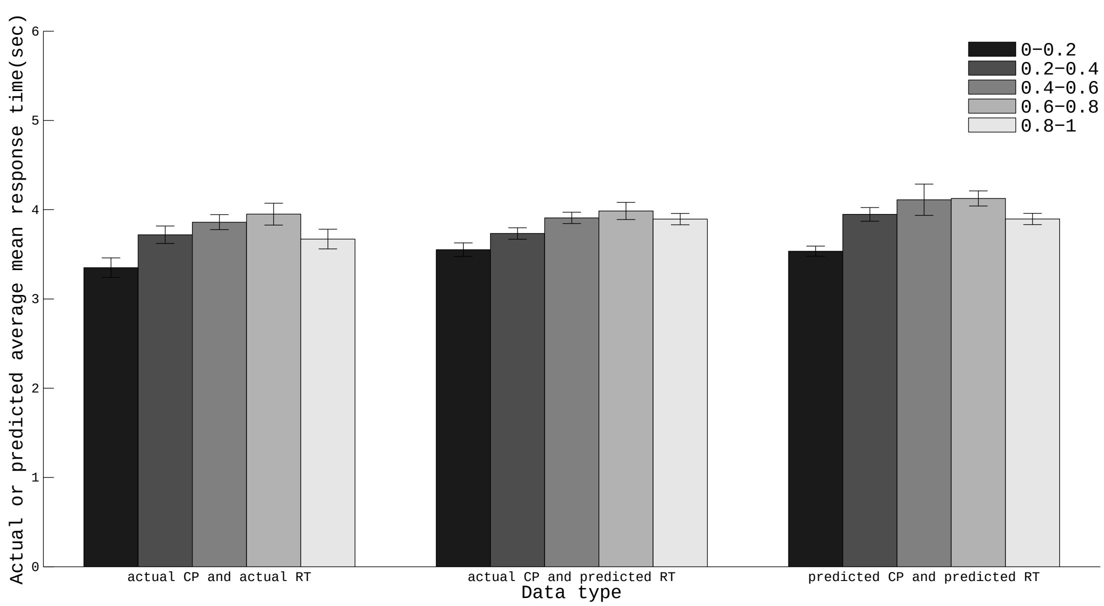

Figure 4. Marginal relationships between choice proportions and response times in Experiment 1. Each bar is associated with a specific range of choice proportions of the LL options. Error bars show 95% confidence intervals.

Figure 5 shows the actual and predicted conditional relationships between choice proportions and response times. It is readily seen that options chosen more frequently tended to have shorter response times than those chosen less frequently ($M_{more} = 3.88s$, $M_{less} = 4.40s$, $t[58] = -3.49$, $p < .01$). The same pattern was also found in individual data of 38 out of 61 participants, with 7 statistically significant differences.

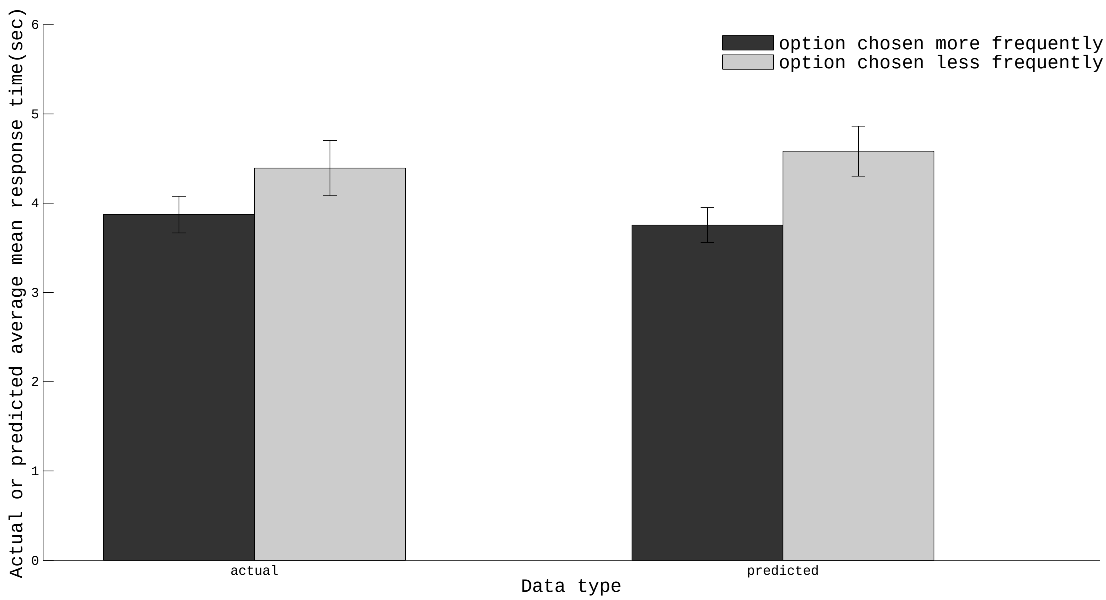

Figure 5. Conditional relationships between choice proportions and response times in Experiment 1. Error bars show 95% confidence intervals.

Preliminary model fitting and comparisons. As in previous study, I first fit all the probabilistic models to individual choice data for a preliminary evaluation of model performance. Table 4 lists the three best models in terms of overall AIC and BIC values. Again, most of the best models were based on an attribute-wise core theory with direct differences. When overall AIC value was used as a criterion, all the best models had a dynamic structure. To the contrary, certain static models matched up with dynamic ones in terms of overall BIC values. For those participants who always chose between immediate SS options and delayed LL options, all of the three best models were attribute-wise ones built upon direct differences, no matter whether overall AIC or BIC value across participants was used as a selection criterion.

Table 4

The Three Best Models for Choice Data in Experiment 1 in Terms of Overall AIC Value or Overall BIC Value across Participants

Number	Transformations of objective value and time	Core theory	Stochastic specification	Number of free parameters	Overall criterion value
			AIC		
1	Power with exponents no larger than 2	Attribute-wise with mixed direct differences	Diffusion model with varied σ, fixed θ^* and varied z^*	6	8086
2	Power with exponents no larger than 2	Attribute-wise with mixed direct differences	Diffusion model with varied σ, fixed θ^* and no initial bias	5	8161
3	Power with exponents no larger than 2 and a scaling constant on time	Alternative-wise with exponential discount function	Diffusion model with varied σ, fixed θ^* and fixed z^*	7	8394
			BIC		
1	Power with exponents no larger than 2	Attribute-wise with single direct differences	Logistic model	4	10437
2	Power with exponents no larger than 2	Attribute-wise with mixed direct differences	Diffusion model with varied σ, fixed θ^* and no initial bias	5	10448
3	Power with exponents no larger than 2	Attribute-wise with single direct differences	Probit model with fixed σ	4	10480

Test of the common difference effect. The test of the common difference effect among participants whose questions involved both immediate and delayed SS options led to some interesting results. According to the common difference effect in intertemporal choice, when both options are further delayed by the same duration, the LL option would become more attractive. Consequently, the choice proportion of the LL option should increase from a probabilistic perspective on intertemporal choice. The actual data, however, suggested that 6 out of the 31 participants who encountered both immediate and delayed SS options shifted their preference toward the SS options when the SS options were delayed options (i.e., in 10 days) as opposed to immediate ones. By contrast, only 1 out of the 31 participants showed a clear choice pattern that was consistent with the common difference effect. The change in choice proportion of the LL options for each of the remaining 24 participants was not statistically significant. Hereafter, I will call the empirical pattern opposite to that suggested by the common difference effect the reversal of common difference effect.

Discussion

The results of this study replicated those from Experiment 3 in Dai and Busemeyer (2014) in the sense that they provided converging evidence for the probabilistic, dynamic, and attribute-wise perspective on intertemporal choice. First, the probabilistic choice behavior revealed in the data from each participant again suggested that intertemporal choice is not deterministic as suggested by the traditional perspective on this topic. Second, the same marginal and conditional relationships between actual choice proportions and mean response times were found as in previous studies. These findings implied again that intertemporal choice is a dynamic process which requires different amounts of deliberation time for different pairs of options. Consequently, it is desirable to develop dynamic instead of static models for intertemporal choice since only the former might be able to account for the empirical regularities with regard to response time data.

The result of preliminary model comparisons also corroborated the previous finding that people tended to use an attribute-wise approach to intertemporal choice. Participants in this experiment were required to indicate their preference between immediate SS options and delayed LL options on half or all of the trials. An alternative-wise choice strategy seemed more efficient in this situation because only one discounted utility needed to be calculated in each question. By contrast, when both options in a question were delayed ones, the traditional delay discounting paradigm suggested that people needed to figure out the discounted utilities twice to make a choice. The increased mental workload made it less likely to adopt an alternative-wise strategy in the latter case. Therefore it should not be surprising that when almost all the questions involved delayed SS options as in Dai and Busemeyer (2014), an attribute-wise approach may be preferred to an alternative-wise one. The results of the current study, however, suggested that an attribute-wise approach to intertemporal choice may be applied more generally to cover the typical question form used in traditional studies on the topic. Most of the best models in this experiment were attribute-wise ones, and this was still true even when only the data from participant who always chose between immediate SS options and delayed LL options were considered. In summary, an attribute-wise approach to intertemporal choice is more likely to be adopted than an alternative-wise one no matter whether the choice questions involve immediate or delays SS options.

Finally, the somewhat unexpected result from testing the common difference effect suggested that delays might be perceived in a way different from what a typical psychophysical function would imply. For most physical stimuli, the psychophysical function mapping objective magnitude to perceived subjective strength assumes a concave form and thus entails decreasing sensitivity to increments in objective magnitude (Stevens, 1957). If this is also true for the perception of time delays, then the common difference effect can be conveniently explained by the relevant psychophysical function since it implies that

$$p(t) - p(0) > p(t + s) - p(s). \tag{38}$$

The reversal of the common difference effect revealed in the empirical data from a handful of participants, however, suggests that the psychophysical function for time delay might be convex at least around certain delay durations. Take the sample questions in Table 3 as an example. It is possible that, in this specific context, participants perceived a delay shorter than 20 days as acceptable but a delay longer than 20 days much more unpleasant. As a result, when one option took place in less than 20 days but the other occurred in more than 20 days, participants might choose the SS option to avoid the longer delay. In general, this could be captured by a power transformation function of objective time with an exponent larger than 1. This was the major reason that I tried two types of power transformation functions of objective time with different ranges of exponents. The same might also be true for the perception of reward in an intertemporal choice scenario. Therefore, I also tried two types of power transformation functions of objective value with different ranges of exponents. The result of model comparison suggested that both time and money might be perceived in a way different from that suggested by a concave or linear transformation function.

Experiment 2

Method

Forty-seven participants (30 females and 17 males) with an average age of about 20 were recruited for this experiment. On average participants received about 12 dollars in 37 days. Seven participants produced extreme choice responses, i.e., always choosing the LL options or the SS options across the whole stimulus set. As a result, their data were excluded from further analysis. Among the remaining 40 participants, there were 28 females and 12 males with the same average age as the original sample. The average amount and delay of payment were virtually the same as the original sample.

In this study, the SS options in all of the questions were delayed ones and the questions were designed to test the transitivity of intertemporal preference in terms of WST. As in Experiment 1, a double-limit adjustment procedure was implemented for each participant at the beginning of the study right after the practice session to find an approximately indifferent pair of intertemporal options. The delay of the SS option in the approximately indifferent pair was always 5 weeks and that for the LL option was always 6 weeks. Eight options were then derived from the two options in the approximately indifferent pair. The delays of the options ranged from 1 week to 10 weeks (excluding 5 and 6 weeks) and the associated reward amounts increased in a linear way. After that, sixteen pairs of options were created to set up three sets of choice questions that were appropriate for examining WST. Specifically, the first set involved options with delays of 1, 2, 3, and 4 weeks respectively; the second set involved options with delays of 7, 8, 9, and 10 weeks respectively; and the last set involved options with delays of 1, 4, 7, and 10 weeks respectively. The first set had relatively short intervals and short delays, the second had relatively short intervals but long delays, and the last had relatively long intervals. All possible pairs from each set of 4 options were then used to generate 6 formal choice questions. In total, there were 16 unique questions for each participant, and each question was presented 10 times. Note that the number of repetitions was doubled compared to that in the previous studies to obtain a more precise estimation of choice probability for each question. This would in turn facilitate detection of intransitivity of intertemporal preference, if any. See Table 5 for the 16 unique questions a typical participant encountered in the study.

Table 5

Formal Questions for a Typical Participant in Experiment 2

Question	Smaller reward (dollars)	Shorter delay (weeks)	Larger reward (dollars)	Longer delay (weeks)
1	10	1	15	2
2	10	1	20	3
3	10	1	25	4
4	15	2	20	3
5	15	2	25	4
6	20	3	25	4
7	40	7	45	8
8	40	7	50	9
9	40	7	55	10
10	45	8	50	9
11	45	8	55	10
12	50	9	55	10
13	10	1	40	7
14	10	1	55	10
15	25	4	40	7
16	25	4	55	10

In addition to the analyses performed across all three studies, a Bayesian model comparison was conducted in this study to investigate the issue of intransitive intertemporal preference. Specifically, WST implies certain constraints among the choice probabilities of the LL options for the 16 unique questions each participant encountered in the study. Take the first six questions in Table 5 as an example. Let p_1, p_2, p_3, p_4, p_5 and p_6 represent the choice probabilities of the LL options in these questions respectively. WST entails that

$$p_1 \geq .5, p_4 \geq .5 \Longrightarrow p_2 \geq .5, p_1 \leq .5, p_4 \leq .5 \Longrightarrow p_2 \leq .5, \tag{39}$$

$$p_1 \geq .5, p_5 \geq .5 \Longrightarrow p_3 \geq .5, p_1 \leq .5, p_5 \leq .5 \Longrightarrow p_3 \leq .5, \tag{40}$$

$$p_2 \geq .5, p_6 \geq .5 \Longrightarrow p_3 \geq .5, p_2 \leq .5, p_6 \leq .5 \Longrightarrow p_3 \leq .5, \tag{41}$$

$$p_4 \geq .5, p_6 \geq .5 \Longrightarrow p_5 \geq .5, p_4 \leq .5, p_6 \leq .5 \Longrightarrow p_5 \leq .5. \tag{42}$$

The same was true for the other sets of questions. From a Bayesian perspective, the constraints imposed by WST can be represented by a constrained model with a joint prior distribution that fulfills these constraints (Cavagnaro & Davis-Stober, in press). Furthermore, to examine whether WST holds in a particular dataset, we can compare the constrained model with a more general model that allows for any relationships among the 16 choice probabilities. Specifically, we need to calculate the marginal likelihood of the data given each model and the resultant Bayes factor between the constrained and more general models for the dataset to select a better model. The Bayesian approach to model comparison assumes that a model is defined by not only the likelihood of the data given certain parameter values but also the prior belief on all possible parameter values in terms of their credibility. Technically speaking, the former is represented as the likelihood function $P(D|\theta, M)$, in which D denotes the data and θ represents the parameter(s) involved in a Model M. And the latter is characterized by $p(\theta|M)$, the prior probability density (or in some cases, probability mass) function of the parameter θ of the model. Here the word prior means that the relevant belief on the credibility of possible parameter values exists prior to collecting data or incorporating the information in the data to update our belief. Similarly, the belief on parameter values after updating with the data is called posterior belief, represented as $\mathrm{P}(\theta|D, M)$. The marginal likelihood of the data given a model is defined as

$$P(D|M) = \int P(D|\theta, M)P(\theta|M)d\theta. \tag{43}$$

We can interpret $P(D|M)$ as a measure of the overall performance of the model with regard to the data, because all possible parameter values of the model are taken into account and their prior credibilities serve as the weights in calculating the marginal likelihood of the data. With the marginal likelihood, we can compare two models, M_1 and M_2, by examining the ratio of marginal likelihood between them. The ratio is actually the Bayes factor (BF) mentioned above, i.e.,

$$BF_{12} = \frac{P(D|M_1)}{P(D|M_2)}. \tag{44}$$

Because the marginal likelihood of data given a model is a measure of overall performance of the model, a value of BF_{12} greater than 1 suggests that Model 1 performs better than Model 2, and vice versa.

In practice, it could be difficult to find the Bayes factor because the precise value of marginal likelihood of each model involves an integral that usually lacks an analytical solution. Fortunately, for constrained models like the one implied by WST, statisticians have devised a convenient way to calculate the Bayes factor (Klugkist & Hoijtink, 2007). Specifically, let $1/c_t$ be the proportion of the prior distribution of the more general model in agreement with the constraints imposed by WST, and $1/d_t$ be the proportion of the posterior distribution of the more general model in agreement with the constraints. It has been shown that the Bayes factor between the constrained model (i.e., the one implied by WST) and the more general model equals

$$BF_{CG} = \frac{p(D|M_C)}{p(D|M_G)} = \frac{c_t}{d_t}. \tag{45}$$

In other words, to get the required Bayes factor, we only need to update the prior distribution of the more general model with the data to find its posterior distribution and then count the relevant proportions. As suggested by Cavagnaro and Davis-Stober (in press), I assumed that the prior distribution of the more general model was uniform in the unit hypercube, $[0,1]^{16}$ to further simplify the problem. This setting facilitated the calculation of the Bayes factor in a number of aspects. First, under this setting, the proportion of the prior distribution of the more general model in agreement with the constraints imposed by WST, i.e., $1/c_t$, was simply the volume of the parameter space of the constrained model (Cavagnaro and Davis-Stober, in press). Therefore, it was possible to use simple algorithms such as a rejection sampling Monte Carlo algorithm to get the value of $1/c_t$. Second, this setting entailed that the 16 parameters were independent of one another in the prior and thus made it possible to obtain analytical solution to the posterior distribution. Specifically, the multidimensional posterior distribution was a product of independent beta distributions of the relevant parameters, and the parameters for each independent beta distribution could be conveniently calculated from the number of times the LL option was chosen out of the 10 trials. For example, if a participant chose the LL option in a particular question 6 times, then the relevant posterior beta distribution was simply $beta(1+6, 1+10-6)$, i.e., $beta(7,5)$. Finally, the resultant analytical solution in turn made it easy to sample from the posterior distribution to count the required proportion, i.e., $1/d_t$. With the values on c_t and d_t, it was straightforward to calculate the Bayes factor with Equation 45. In addition, Klugkist and Hoijtink (2007) showed that the Bayes factor calculated using the current method is relatively robust to different choices of prior distribution as long as certain quite general conditions are satisfied. Therefore, it was appropriate to use the current prior as a convenient way to calculate the Bayes factor. In this study, c_t was estimated based on 10,000,000 samples from the prior distribution of the more general model, and d_t was estimated based on 100,000 samples from the corresponding posterior distribution. Because each participant answered every unique intertemporal choice question multiple times in this study, I conducted the Bayesian analysis on individual data instead of running an aggregate analysis to test WST in intertemporal choice.

Results

Choice variability as an indication of probabilistic nature of intertemporal choice. As expected, each participant in this study switched between the SS and LL options in multiple unique questions when these questions were presented repeatedly. On average, participants changed their mind in about 10 out 16 unique questions.

Relationships between choice proportions and response times. Figure 6 demonstrates the actual and predicted marginal relationships between choice proportions and response times in the current study. As before, an inverse U-shaped relationship between actual choice proportions and mean response times was revealed across questions and participants ($M_{extreme} = 3.91s$, $M_{moderate} = 4.34s$, $t[35] = -4.43$, $p < .01$). The same pattern was found in individual data from 28 out of the 40 participants, with 6 differences reaching statistical significance. [5]

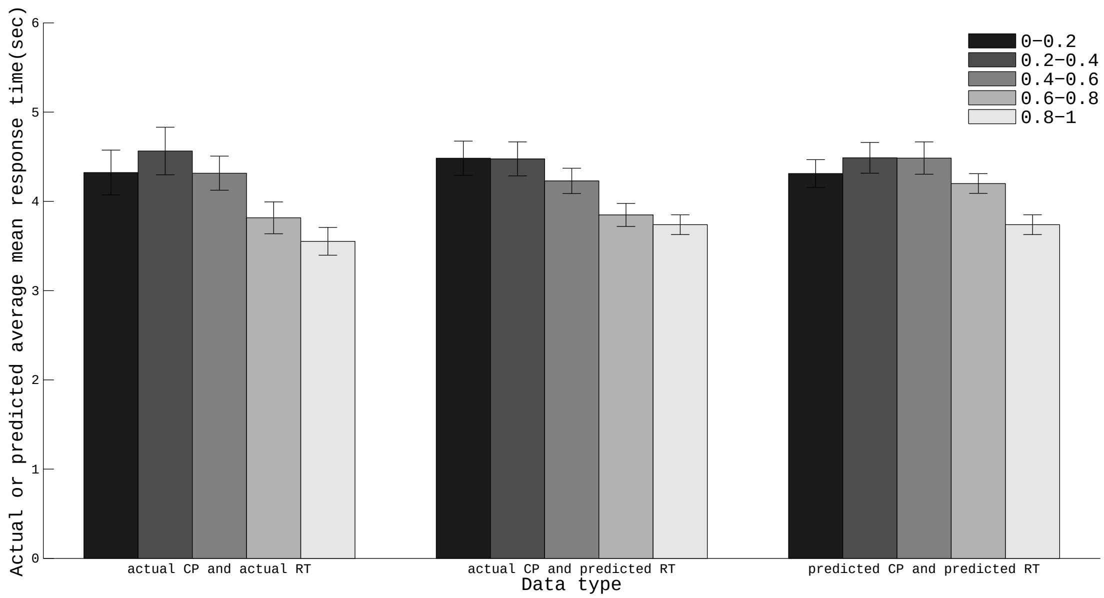

Figure 6. Marginal relationships between choice proportions and response times in Experiment 2. Each bar is associated with a specific range of choice proportions of the LL options. Error bars show 95% confidence intervals.

Figure 7 shows the actual and predicted conditional relationships between choice proportions and response times. Again options chosen less frequently tended to associate with longer response times than those chosen more frequently ($M_{more} = 4.08s$, $M_{less} = 5.16s$, $t[39] = -8.59$, $p < .01$). The same pattern was also found in individual data from 37 out of the sample of 40 participants, with 15 reaching statistically significance.

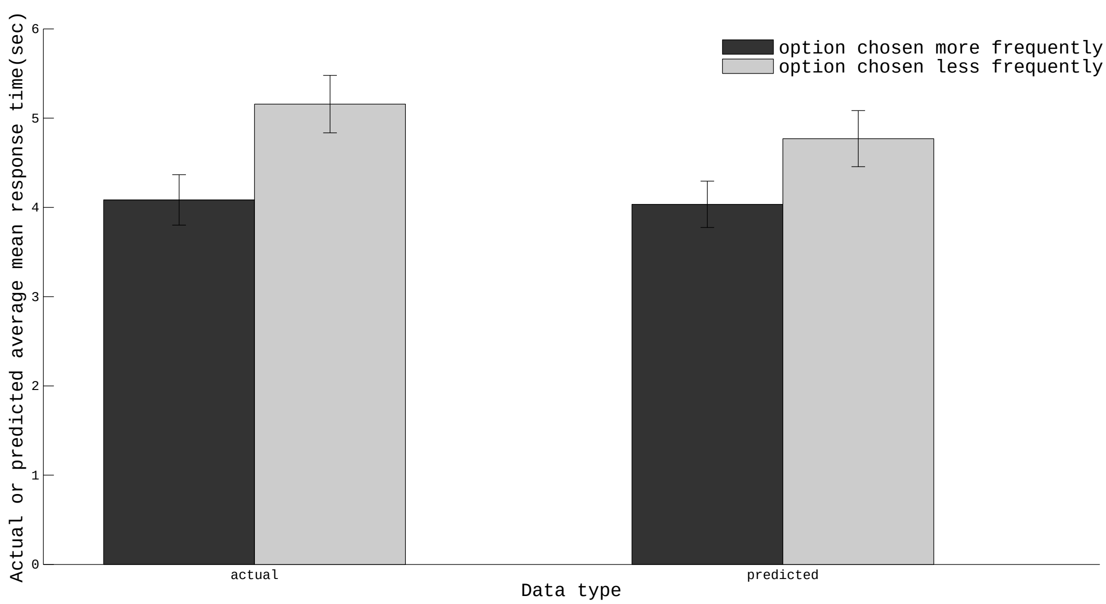

Figure 7. Conditional relationships between choice proportions and response times in Experiment 2. Error bars show 95% confidence intervals.

Preliminary model fitting and comparisons. As in previous studies, I first fit all the probabilistic models to individual choice data for a preliminary evaluation of model performance. Table 5 lists the three best models for choice data in terms of overall AIC and BIC values. A couple of patterns can be found from the table. First, all the best models were attribute-wise ones based on either direct or relative differences. Second, whether a model with direct or relative differences performed the best depended on the specific criterion used, i.e., overall AIC or BIC value across participants. Similarly, the two criteria also favored different models with regard to model dynamic property, i.e., whether a model assumes a static or dynamic stochastic specification. Specifically, when using overall AIC value, the best two models were both dynamic ones, followed by a static tradeoff model. By contrast, when BIC was of concern, the best two models were static ones, followed by a dynamic diffusion model. This was similar to those found in previous studies.

Transitivity in intertemporal preference. Table 6 lists the Bayes factor between the constrained model and the more general model for each participant of the current study and Figure 8 shows the distribution of log(Bayes factor). It can been seen from the histogram that most participants (33 out of 40) had a Bayes factor greater than 1 (i.e., log(BF) > 0), suggesting that the constrained model described the data better than the more general model for these participants. The sum of log(BF) across all participants was 59.66, decisively supporting the constrained model implied by WST. Table 7 shows the choice proportions of the LL options for questions answered by a typical participant whose data suggested a satisfaction of WST (subject 5), those for a typical participant whose data suggested a violation of WST (subject 13), and the average results across participants.

Table 5

The Three Best Models for Choice Data in Experiment 2 in Terms of Overall AIC Value or Overall BIC Value across Participants

Number	Transformations of objective value and time	Core theory	Stochastic specification	Number of free parameters	Overall criterion value
			AIC		
1	Power with exponents no larger than 2	Attribute-wise with mixed direct differences	Diffusion model with varied σ, fixed θ^* and varied z^*	6	4231
2	Power with exponents no larger than 2	Attribute-wise with mixed direct differences	Diffusion model with varied σ, fixed θ^* and no initial bias	5	4264
3	Logarithm	Attribute-wise with single direct differences	Tradeoff model with ratio rule	6	4387
			BIC		
1	Identity	Attribute-wise with relative differences	Proportional difference model	2	5267
2	Identity	Attribute-wise with relative differences	Random preference model	2	5432
3	Power with exponents no larger than 2	Attribute-wise with mixed direct differences	Diffusion model with varied σ, fixed θ^* and no initial bias	5	5593

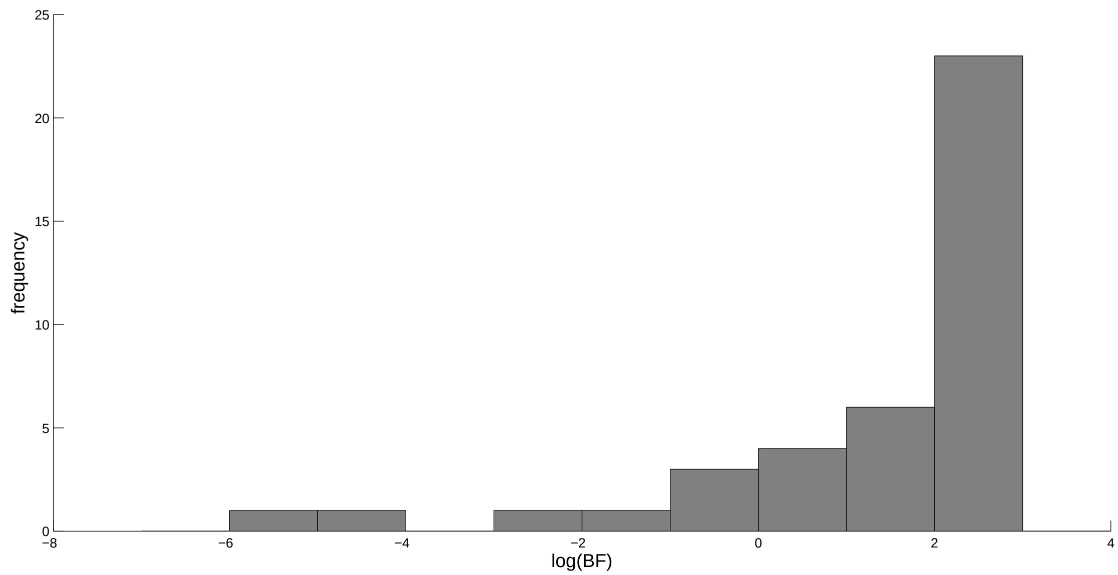

Figure 8. Distribution of log(Bayes factor) for the test of transitivity of intertemporal preference.

Table 6

Individual Bayes Factors for Transitivity of Intertemporal Preference with Regard to Weak Stochastic Transitivity

Participant	1	2	3	4	5	6	7	8	9	10
Bayes factor	4.01	6.18	7.21	18.88	8.21	17.49	1.26	17.71	18.87	10.11
Participant	11	12	13	14	15	16	17	18	19	20
Bayes factor	0.25	0.01	0.00	18.68	13.72	18.39	6.54	4.56	14.81	16.41
Participant	21	22	23	24	25	26	27	28	29	30
Bayes factor	15.81	1.90	17.89	12.16	16.32	5.91	9.15	0.86	1.13	17.34
Participant	31	32	33	34	35	36	37	38	39	40
Bayes factor	0.09	2.58	0.40	17.20	14.82	18.20	15.59	16.37	18.61	0.63

Table 7

Choice Proportions of the LL Options for Typical Individual Data either Satisfying (Subject 5) or Violating (Subject 13) WST, and the Average Result Across Participants

Question	Shorter delay (weeks)	Longer delay (weeks)	Choice proportion of the LL option for Subject 5	Choice proportion of the LL option for subject 13	Average Choice proportion of the LL option
1	1	2	0.4	0.9	0.38
2	1	3	0	0	0.39
3	1	4	0	0	0.36
4	2	3	0.7	0.8	0.53
5	2	4	0.6	0	0.50
6	3	4	1	0.9	0.58
7	7	8	0.9	0.9	0.62
8	7	9	0.7	0.1	0.63
9	7	10	0.7	0	0.67
10	8	9	0.9	0.9	0.68
11	8	10	0.9	0	0.67
12	9	10	1	0.8	0.70
13	1	7	0.1	0	0.38
14	1	10	0.1	0	0.46
15	4	7	0	0	0.52
16	4	10	0	0	0.54

Discussion

The results of this experiment were generally the same as those from previous two studies. First, the probabilistic nature of intertemporal choice was again supported by the data since all participants changed their mind in multiple unique questions asked repeatedly. All the existing evidence suggested that intertemporal choice is essentially probabilistic just like risky choice. Consequently, researchers interested in this topic should try to develop relevant models that could accommodate this property. This was exactly the major motive for the current line of research. Second, each of the best models in terms of either overall AIC or overall BIC value across participants implied an attribute-wise decision strategy for intertemporal choice, which is fundamentally different from that of the traditional delay discounting paradigm with an alternative-wise perspective. This result again suggested that it is the time to abandon the popular but psychologically unlikely delay discounting paradigm for intertemporal choice, at least for the task of binary intertemporal choice between gains that has been substantially studied in this line of research. Third, the two model selection criteria, i.e., overall AIC and BIC values across participants, tended to pick different attribute-wise models in terms of the type of differences involved and model dynamic property. It seems appropriate to postpone the decision on the best model to a later section where an overall model comparison was conducted with regard to the data across all three studies and using more selection criteria.

Another important finding in the current study was that intertemporal choice tended to be transitive at an individual level with regard to WST. This result echoed recent progress in the literature suggesting that risky preference is actually transitive at an individual level(Regenwetter, Dana, & Davis-Stober, 2011; Regenwetter & Davis-Stober, 2012; Cavagnaro & Davis-Stober, in press) However, the current result was inconsistent with that from a previous study on the same issue (Roelofsma & Read, 2000). The major reason for the incongruent results was that the current study used different ways to examine the issue of transitivity. First of all, this study was aimed at studying the transitivity of intertemporal preference at an individual level while the previous one analyzed the aggregate data across participants to make a conclusion. The most prominent problem with the latter method lies in the implicit assumption that all participants use the same strategy to make intertemporal choice and have the same parameter values on the corresponding model. This assumption is clearly untenable given the large amount of literature on individual differences in intertemporal choice. For example, in the early days before the delay discounting paradigm gained dominance in the research area, a number of scholars had recognized variations in intertemporal choice behavior and attributed them to different factors, such as individual differences in ability to imagine the future as well as the degree of psychological discomfort associated with self-denial (Frederick et al., 2002). Everyday experience also suggests that people differ substantially in terms of degree of impulsivity: Some people might want to wait a long time for a small increase in reward while others simply grasp any immediate gains regardless how much they have to give up in the future for their choice at the moment. Given all kinds of individual differences that might impact intertemporal choice, it is possible that individual data are in agreement with stochastic transitivity defined by WST while the aggregate data suggest the opposite situation. Imagine that three people A, B, and C have different preferences among three intertemporal options X, Y, and Z, with $X \succ Y \succ Z$ for A, $Y \succ Z \succ X$ for B, and $Z \succ X \succ Y$ for C. This suggests that the preference relation among the three options is transitive and deterministic for each individual. When we look at the aggregate situation, however, a probabilistic and intransitive pattern arises. Specifically, two-thirds of the people (i.e., A and C) prefer X to Y, two-thirds of the people (i.e., A and B) prefer Y to Z, but again two-thirds of the people (i.e., B and C) prefer Z to X. In other words, deterministic and transitive individual data can produce probabilistic and intransitive patterns when they are analyzed as a whole. Therefore, it is problematic to analyze aggregate data for a conclusion on the transitivity of intertemporal preference. The current example also provides a rationale for analyzing individual data to explore the probabilistic nature of intertemporal choice.

Besides the aforementioned difference in research and analysis methods between the current and previous studies, the current study adopted a state-of-the-art approach to testing WST, i.e., the Bayesian model comparison method, while the previous study relied on the observed ratio of different intransitive patterns to make a conclusion. This Bayesian model comparison approach has two major advantages over the method used in the previous study. First, we can use the Bayesian approach without invoking any specific model of intertemporal choice, but for the method used previously, certain delay discounting model and stochastic specification are necessary for determining the cutoff value on the observed ratio, based on which a conclusion on intransitivity can be made. In other words, the former approach is more general than the latter in terms of prerequisites for their application. Second, the Bayesian approach can be used on individual data while the method in the previous study is only applicable to aggregate data. Therefore, the latter is of less use than the former, especially considering the problems with analyzing the aggregate data in this case. In summary, the current approach to analyzing transitivity of intertemporal preference is more sophisticated and general and thus the resultant conclusion deserves more confidence than that from the previous study.

It is worth noting that, although the majority of participants showed transitive intertemporal preference, there were still a small portion of participants whose data suggested the opposite even according to the weakest definition of stochastic transitivity, i.e., WST. Consequently, if we need a comprehensive model that can not only perform the best when all individual data were considered but also accommodate individual differences in revealed qualitative results, then the selected model should be able to predict both transitive and intransitive intertemporal preferences depending on the specific values of model parameters.

An Overall Model Comparison and Model Prediction

For each of the studies discussed above, I've reported the results of preliminary model comparisons using overall AIC and BIC values across participants as the selection criteria. In general these results were in favor of a dynamic and attribute-wise perspective on intertemporal choice as found in Dai and Busemeyer (2014). The best model, however, varied across studies and depended on model selection criterion. In what follows, I will report the results of an overall model comparison across all the studies covered in this dissertation using different model comparison methods. The purpose of this analysis is to find an overall best model among all the competing static and dynamic models of intertemporal choice examined in this dissertation. The winning model should not only provide a general framework for various effects and phenomena in intertemporal choice but also lead to the best quantitative performance in general when data from all the studies were taken into account.

Method

Model comparisons. Given the large number of models examined in this dissertation, it was necessary to first pick a small set of candidate models before applying various model selection methods to determine the overall best model. To this end, I used overall AIC and BIC values across participants with regard to individual choice data as before but considered data from all the studies covered in this dissertation. To get a reasonable number of models for further comparison, I retained in the candidate set the three best models for individual choice data in terms of either overall AIC value or overall BIC value across participants.

After determining the small set of candidate models, I proceeded to examine their performance on individual choice data in terms of counts of lowest AIC or BIC values and the results of pairwise comparisons based on AIC and BIC values. Although these and previous methods were all built upon AIC and BIC values, the current ones provided new information from different angles for model selection. In addition, I tried a couple of other model comparison methods that address the issue of model complexity more properly and comprehensively than AIC and BIC. Specifically, for each participant, I ran a cross-validation test and a Bayesian model comparison on the candidate models. The cross-validation test was first proposed by Mosier (1951) and it involves first dividing the data set into two samples and then using one sample for calibration and the other for validation. Because complicated models tend to overfit the calibration data by accommodating more noise, they are prone to fit the validation sample more poorly than the true but simpler model. This is the essential way the cross-validation test balances between model complexity and goodness-of-fit. To generate an overall index for model selection, I ran the cross-validation test for each participant and then summed up the results across participants. For each participant I randomly divided his/her data into two equal-size samples, fit candidate models to one sample (i.e., the calibration sample), and then used the parameter estimates from the calibration sample to predict the likelihood of the data in the validation sample as a measure of model performance. This procedure was implemented fifty times and the results were averaged to get a reliable measure for each model.

Yet another approach to considering model complexity when comparing different models is to run a full Bayesian analysis and then use the Bayes factor to select the best model. This approach takes into account not only number of parameters involved in a specific model but also the functional forms and the extension of the parameter space(Myung & Pitt, 1997) . Therefore, it incorporates more aspects of model complexity that might impact the goodness-of-fit of a model than other methods. As mentioned above, the Bayesian approach to model comparison assumes that a model is defined by not only the likelihood of the data given certain parameter values but also the prior belief on all possible parameter values in terms of their credibility. When only the likelihood function $P(D|\theta, M)$ is considered as the essential part of a model, the issue of model comparison is pivoted around the maximum likelihood of data given each of the models to compare. To the contrary, when both likelihood function and prior belief of the parameter(s) are taken into account, the pivotal issue becomes finding a model that produces the highest average likelihood of the data across the whole parameter space. In other words, if a model can produce the highest maximum-likelihood of the data among competing models but most of its possible parameter values lead to quite low likelihood of the data, then it should not be selected as the best model because on average it does not fit the data well. Obviously model complexity plays a critical role in determining the overall performance of a model. For example, a model with a wider range of parameter values produces at least the same maximum-likelihood of the data as a model with a narrower parameter space, but the extra range of parameter values of the former may produce quite low likelihood of the data and thus worsen the overall performance of the model. In addition, the extra range of parameter values would also attenuate the credibility of each parameter value and thus counterbalance the impact of a higher maximum-likelihood on the overall performance of the model.

As mentioned above, it could be difficult in practice to find the Bayes factor because the precise value of marginal likelihood of each model involves an integral that usually lacks an analytical solution. One "brutal force" solution to this problem is to use grid approximation to calculate each marginal likelihood. Specifically,

$$P(D|M) = \int P(D|\theta, M)P(\theta|M)d\theta \approx \sum_{\theta} P(D|\theta, M)P(\theta|M). \tag{46}$$

The sum in Equation 46 is over discrete (and usually equally-spaced) values of θ and $P(\theta|M)$ in the sum represents a probability mass at each value of θ(Kruschke, 2010). In this dissertation, I used an uninformative prior distribution for each candidate model by setting a uniform distribution within a reasonable range for each parameter. The parameter space for each model was then partitioned into many subspaces with equal probability mass and the likelihood of the data given the parameter value at the center of each subspace was used in the right-hand side of Equation 46 to estimate marginal likelihood. For each model, at least 6,000,000 subspaces were used to estimate the marginal likelihood of choice data from each participant.

Finally, I fit the dynamic models in the limited set to individual choice and response time data simultaneously (using Equation 33) and ran a model comparison using AIC and

BIC values. In this way, more information from the empirical data was employed to distinguish among competing diffusion models. Based on the results of all the model comparisons, a single best model was selected.

Model predictions. Yet another way of examining the fitting performance of a specific model is to compare its predictions with the actual data. If a model captures the key components of the actual data-generating mechanism, then its predictions based on the best-fitting parameters should match the observed data closely. Three types of information in the data are of particular interest in this dissertation, i.e., the choice proportion of the LL option for each repeated presented question, the mean response time for each question, and the relationships between choice proportions and response times. Therefore, after selecting an overall best model based on all the aforementioned methods and across all participants, I used each individual's best-fitting parameters for both choice and response time data to predict the choice probability of the LL option and the conditional mean response time given the actual choice on each trial. For each question, I then averaged the predicted results on all the trials in which the questions was presented to get predicted choice probability of the LL option (the same as in each repeated trial) and predicted mean response time (i.e., the predicted marginal response time). With the predicted choice probabilities and mean response times for all questions across participants, two correlational analyses were performed to compare the actual choice proportions with the predicted choice probabilities and to compare the actual mean response times with the predicted mean response times.

In addition, I examined the marginal relationship between actual choice proportions and predicted mean response times, the marginal relationship between predicted choice probabilities and predicted mean response times, and the conditional relationship between predicted choice probabilities and predicted mean response times to see whether the best model predicted the same patterns as revealed in the actual data. The latter two tests were absent in Dai and Busemeyer (2014) but they provided additional ways to show the validity of the best model. The marginal analysis was built upon actual or predicted choice proportion and mean response time for each question as in the correlational analysis mentioned above. The conditional analysis involved categorizing options in each question as more or less frequently chosen option according to the prediction of the best model and then comparing the predicted mean response times between the two types of options across questions and participants. The predicted results were demonstrated graphically together with the actual ones to show the match between observed and predicted patterns.

Finally, to show that the winning model can predict the three effects studied in Experiment 3 of Dai and Busemeyer (2014), I compared the actual and predicted choice patterns for questions associated with each of the effects. Specifically, for each participant, all the questions associated with each effect were ordered in terms of their target attribute (i.e., delay duration for the delay duration and common difference effects and reward amount for the magnitude effect; see Table 1 for sample questions) so that the corresponding actual and predicted choice proportions could be used to reveal the choice patterns. After that, the actual and predicted choice patterns for a typical participant whose data revealed all three effects and the average patterns across participants were plotted to show the resemblance between the actual and predicted results.

Results

Candidate models. Table 8 lists the three best models for individual choice data in terms of overall AIC and BIC values across all participants in the studies. When data from all three experiments were considered, all the best models were attribute-wise with direct differences and both criteria suggested the same single best model, i.e., the diffusion model with mixed direct differences, power transformations on objective value and time allowing for both diminishing and increasing marginal sensitivity, varied σ, fixed θ^*, and no initial bias. The second best model in terms of overall AIC value was almost the same as the single best model except for the assumption of varied z^*, and the third best model using the same criterion was a probabilistic version of the tradeoff model proposed by Scholten and Read (2010). When overall BIC values were of concern, the second best model was a static random preference model and the third was the diffusion model favored in Dai and Busemeyer (2014). In total, there are five candidate models to be further compared with various other methods. Note that the overall AIC value for the corresponding best constant error model is 22348, and the overall BIC value for the corresponding best constant error model is 26775. Clearly the performance of constant error models was much worse than that of the candidate models, suggesting the inadequacy of "trembling hand" for explaining the probabilistic nature of intertemporal choice. A comprehensive summary of performance of all the models is available in an Excel file upon request.

Alternative AIC- and BIC-based model selection . Table 9 lists the performance of the five candidate models for individual choice data in terms of count of lowest AIC values or BIC values across all participants. When AIC was of concern, the performance of the two diffusion models with mixed direct differences and the random preference model were about the same; when BIC was considered, the performance of the random preference model was virtually the same as that of the diffusion model with single direct differences. Finally, pairwise comparisons based on AIC values led to circular relationship among the candidate models, while that using BIC values slightly favored the random preference model over the diffusion model with single direct differences.

Cross-validation test. Table 10 lists the predicted log-likelihood from each model averaged across 50 repetitions and summed across participants from all the studies. Clearly the diffusion models with mixed direct differences performed the best among the five candidate models, followed by the tradeoff model and the diffusion model with single direct differences that won the model competition in Dai and Busemeyer (2014). The static random preference model performed the worst in terms of predicted log-likelihood in the cross-validation test. It is worth noting that the best model, i.e., the diffusion model with mixed direct differences and no initial bias, actually won the competition in each repetition in the cross-validation test.

Table 8

The Three Best Models for Choice Data in Terms of Overall AIC Value or Overall BIC Value across Participants from All the Studies

Number	Transformations of objective value and time	Core theory	Stochastic specification	Number of free parameters	Overall criterion value
			AIC		
1	Power with exponents no larger than 2	Attribute-wise with mixed direct differences	Diffusion model with varied σ, fixed $\theta^*$and no initial bias	5	18220
2	Power with exponents no larger than 2	Attribute-wise with mixed direct differences	Diffusion model with varied σ, fixed $\theta^*$and varied z^*	6	18230
3	Logarithm	Attribute-wise with single direct differences	Tradeoff model with ratio rule	6	18859
			BIC		
1	Power with exponents no larger than 2	Attribute-wise with mixed direct differences	Diffusion model with varied σ, fixed θ^* and no initial bias	5	23929
2	Power with exponents no larger than 2	Attribute-wise with single direct differences	Random preference model	4	24038
3	Power with exponents no larger than 2	Attribute-wise with single direct differences	Diffusion model with varied σ, fixed θ^* and no initial bias	4	24050

Table 9

Performance of the Candidate Models for Choice Data in Terms of Count of Lowest AIC Values or BIC Values Across All Participants

Model	Transformations of objective value and time	Core theory	Stochastic specification	Number of parameters	Count of lowest AIC value	Count of lowest BIC value
1	Power with exponents no larger than 2	Attribute-wise with mixed direct differences	Diffusion model with varied σ, fixed θ^* and no initial bias	5	32	27
2	Power with exponents no larger than 2	Attribute-wise with mixed direct differences	Diffusion model with varied σ, fixed θ^* and varied z^*	6	33	18
3	Logarithm	Attribute-wise with single direct differences	Tradeoff model with ratio rule	6	16	7
4	Power with exponents no larger than 2	Attribute-wise with single direct differences	Diffusion model with varied σ, fixed θ^* and no initial bias	4	24	42
5	Power with exponents no larger than 2	Attribute-wise with single direct differences	Random preference model	4	33	44

Table 10

Performance of the Candidate Models on Choice Data in Terms of Predicted Log-likelihood Summed Across All Participants in the Cross-validation Test

Model	Transformations of objective value and time	Core theory	Stochastic specification	Number of parameters	Predicted log-likelihood
1	Power with exponents no larger than 2	Attribute-wise with mixed direct differences	Diffusion model with varied σ, fixed θ^* and no initial bias	5	-4783
2	Power with exponents no larger than 2	Attribute-wise with mixed direct differences	Diffusion model with varied σ, fixed θ^* and varied z^*	6	-4904
3	Logarithm	Attribute-wise with single direct differences	Tradeoff model with ratio rule	6	-5033
4	Power with exponents no larger than 2	Attribute-wise with single direct differences	Diffusion model with varied σ, fixed θ^* and no initial bias	4	-5049
5	Power with exponents no larger than 2	Attribute-wise with single direct differences	Random preference model	4	-5161

Bayesian model comparison. Because the tradeoff model does not assume an attention shifting mechanism and thus is fundamentally different from the other four candidate models, I used the marginal likelihood given the tradeoff model as the denominator to calculate Bayes factors for the other models. Figure 9 shows the distribution of log(BF) for each other model and Table 11 lists the number of participants whose data supported the tradeoff model (i.e., $BF < 1$) or the alternative model (i.e., $BF > 1$). It is readily seen that in general the tradeoff model performed worse than other models in describing the individual data. Most Bayes factors were above 1, in favor of the alternative model against the tradeoff model. For example, when comparing the diffusion model with mixed direct differences and varied initial bias against the tradeoff model, individual data of 13 participants favored the tradeoff model, while those of the remaining 125 participants supported the diffusion model. The sum of log(BF) across all participants suggested that the diffusion model with mixed direct differences and varied initial bias performed the best among the five candidate models. See Appendix for specific prior distributions used for the candidate models.

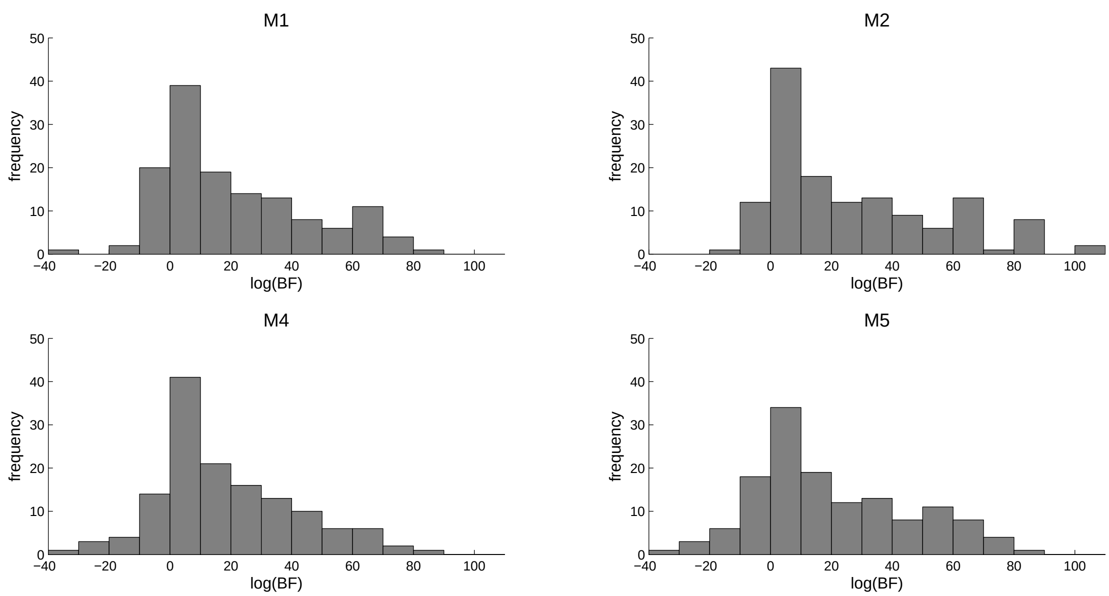

Figure 9. Distribution of Bayes factor for each model against the tradeoff model.

Table 11

Results of Model Comparison Using Bayes Factors

Model	Transformations of objective value and time	Core theory	Stochastic specification	Number of parameters	BF < 1	BF > 1	Sum of log(BF)
1	Power with exponents no larger than 2	Attribute-wise with mixed direct differences	Diffusion model with varied σ, fixed θ^* and no initial bias	5	23	115	2867
2	Power with exponents no larger than 2	Attribute-wise with mixed direct differences	Diffusion model with varied σ, fixed θ^* and varied z^*	6	13	125	3682
4	Power with exponents no larger than 2	Attribute-wise with single direct differences	Diffusion model with varied σ, fixed θ^* and no initial bias	4	22	116	2516
5	Power with exponents no larger than 2	Attribute-wise with single direct differences	Random preference model	4	28	110	2724

Model fitting and comparisons with regard to both choice and response time data. Table 12 lists the overall AIC and BIC values across all participants when the three dynamic models among the candidate models were fit to individual choice and response time data simultaneously. Clearly both model selection criteria favored the diffusion model with mixed direct differences and varied z^*. Table 13 lists the corresponding counts of lowest AIC values and BIC values across all participants. It is readily seen that when AIC was considered, the diffusion model with mixed direct differences and varied z^* again dominated the other two models, especially the model with single direct difference favored in Dai and Busemeyer (2014). By contrast, when BIC was of concern, the two diffusion models with mixed direct differences performed equally well, whereas the diffusion model with single direct differences did not lead to the lowest BIC value for any of the participants. The results of pairwise comparisons were essentially the same. In summary, when both choice and response time data were taken into consideration, the diffusion model with mixed direct differences and varied z^* performed the best among the three dynamic candidate models. Based on this and previous results of model comparison with various methods, it seems appropriate to pick the diffusion model with mixed direct differences and varied z^* as the best model. Table 14 provides a summary of the parameter estimates from the model when it was fit to both choice and response time data.

Table 12

Performance of the Three Candidate Dynamic Models on Both Choice and Response Time Data in Terms of Overall AIC and BIC Values

Model	Transformations of objective value and time	Core theory	Stochastic specification	Number of parameters	Overall AIC value	Overall BIC value
1	Power with exponents no larger than 2	Attribute-wise with mixed direct differences	Diffusion model with varied σ, fixed θ^* and no initial bias	6	121854	128704
2	Power with exponents no larger than 2	Attribute-wise with mixed direct differences	Diffusion model with varied σ, fixed θ^* and varied z^*	7	120382	128374
4	Power with exponents no larger than 2	Attribute-wise with single direct differences	Diffusion model with varied σ, fixed θ^* and no initial bias	5	140053	145763

Table 13

Performance of the Three Candidate Dynamic Models on Both Choice and Response Time Data in Terms of Count of Lowest AIC Values or BIC Values Across All Participants

Model	Transformations of objective value and time	Core theory	Stochastic specification	Number of parameters	Count of lowest AIC values	Count of lowest BIC values
1	Power with exponents no larger than 2	Attribute-wise with mixed direct differences	Diffusion model with varied σ, fixed θ^* and no initial bias	6	50	69
2	Power with exponents no larger than 2	Attribute-wise with mixed direct differences	Diffusion model with varied σ, fixed θ^* and varied z^*	7	88	69
4	Power with exponents no larger than 2	Attribute-wise with single direct differences	Diffusion model with varied σ, fixed θ^* and no initial bias	5	0	0

Table 14

Means and Standard Deviations of Parameter Estimates from the Attribute-wise Diffusion Model with Mixed Directed Differences, Power Transformations of objective Value and Time Allowing for Both Diminishing and Increasing Marginal Sensitivity, Fixed θ^*, Varied σ, and Varied z^*

Parameter	w	θ^*	z^*	α	β	k	Ter
Mean	.54	1.91	.31	1.07	.92	.54	1.34
Standard Deviation	.21	.44	.27	.83	.75	.47	.19

Model predictions. Figure 10 shows the predictions of the best model with regard to Experiment 3 in Dai and Busemeyer (2014) against the actual data across questions and participants. The left panel shows a scatterplot of the predicted choice probabilities of the LL options and the corresponding actual choice proportions, while the right panel shows the relevant result with regard to response times. There were clearly strong associations between the actual and predicted results for both choice proportions and response times. For choice proportions, the correlation coefficient for aggregate data was .90, $p < .01$, and the average correlation coefficient across individual participants was .71. For response times, the correlation coefficient for aggregate data was .76, $p < .01$, with an average correlation coefficient across individual participants of .24. The predicted marginal and conditional relationships between choice proportions and response times could be found in Figures 2 and 3 respectively, which also demonstrated the corresponding actual relationships. Clearly the predicted results showed the same patterns as the observed data.

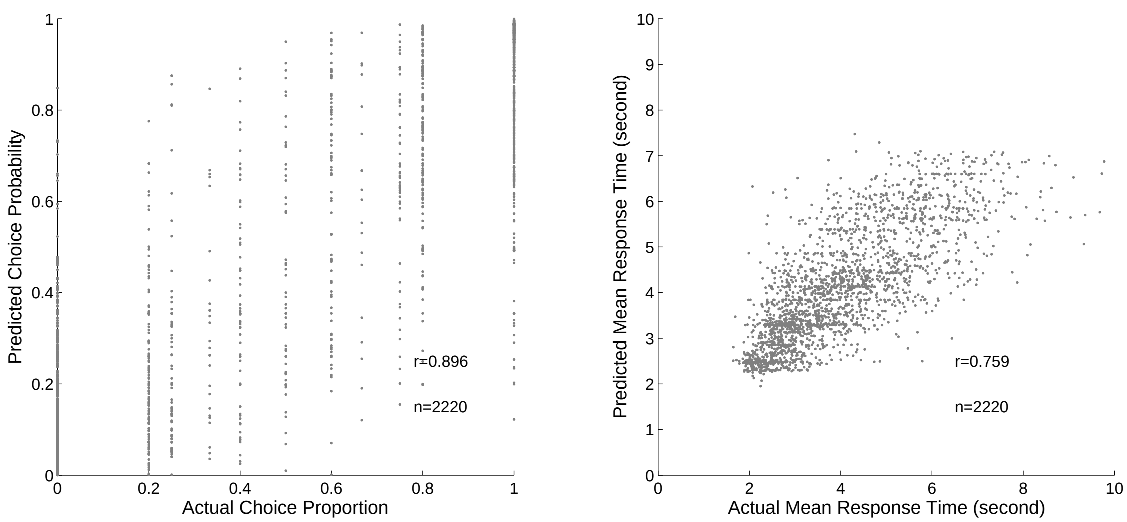

Figure 10. Predictions of the best model for Experiment 3 in Dai and Busemeyer (2014). The left panel shows a scatterplot of the actual choice proportions and the predicted choice

probabilities of the LL options for all questions and participants; the right panel shows a scatterplot of the actual mean response times and the predicted mean response times for all questions and participants.

Similar results were also found for the data of the two new experiments in this dissertation. Figures 11 and 12 show the predictions of the best model with regard to Experiments 1 and 2 respectively against the actual data across questions and participants. The left panels show scatterplots of the predicted choice probabilities of the LL options and the corresponding actual choice proportions, while the right panels show the relevant results with regard to response times. Again there were strong associations between the actual and predicted results for both choice proportions and response times. The aggregate correlation coefficient between actual and predicted choice proportions in Experiment 1 was .89, $p < .01$, with an average correlation coefficient across participants of .78. The aggregate correlation coefficient between actual and predicted mean response times in Experiment 1 was .67, $p < .01$, with an average correlation coefficient across individual participants of .27. The aggregate correlation coefficient between actual and predicted choice proportions in Experiment 2 was .91, $p < .01$, with an average correlation coefficient across participants of .69. The aggregate correlation coefficient between actual and predicted mean response times in Experiment 2 was .77, $p < .01$, with an average correlation coefficient across individual participants of .36. Finally, the predicted marginal and conditional relationships between choice proportions and response times in Experiments 1 and 2, together with the corresponding actual relationships, could be found in Figures 4, 5, 6, 7 respectively. Clearly the predicted results in general showed the same patterns as the observed data.

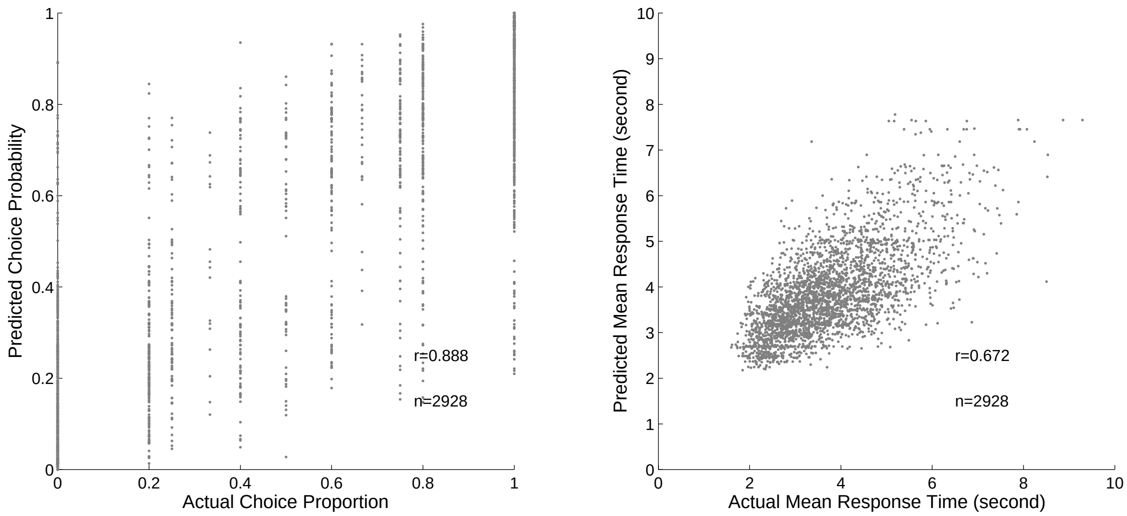

Figure 11. Predictions of the best model for Experiment 1. The left panel shows a scatterplot of the actual choice proportions and the predicted choice probabilities of the LL options for all questions and participants; the right panel shows a scatterplot of the actual mean response times and the predicted mean response times for all questions and participants.

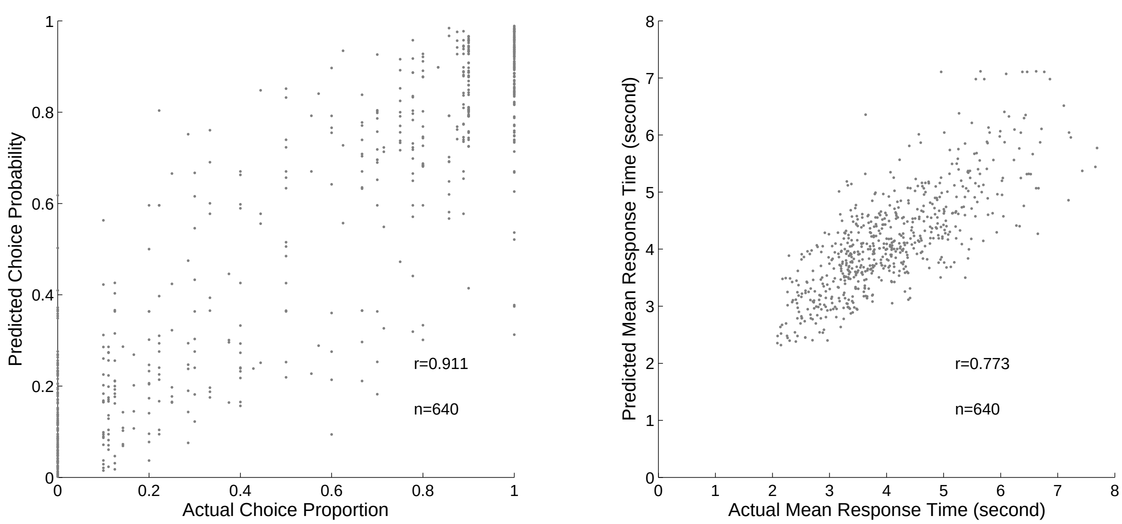

Figure 12. Predictions of the best model for Experiment 2. The left panel shows a scatterplot of the actual choice proportions and the predicted choice probabilities of the LL options for all questions and participants; the right panel shows a scatterplot of the actual mean response times and the predicted mean response times for all questions and participants.

Finally, Figure 13 shows line graphs illustrating the choice patterns of a typical participant and the average results across participants in Experiment 3 of Dai and Busemeyer (2014). The left panels show the actual results and the right panels show the predictions of the winning model. It is readily seen that both individual and average predictions of the winning model captured the general trend of the actual change patterns reasonably well.

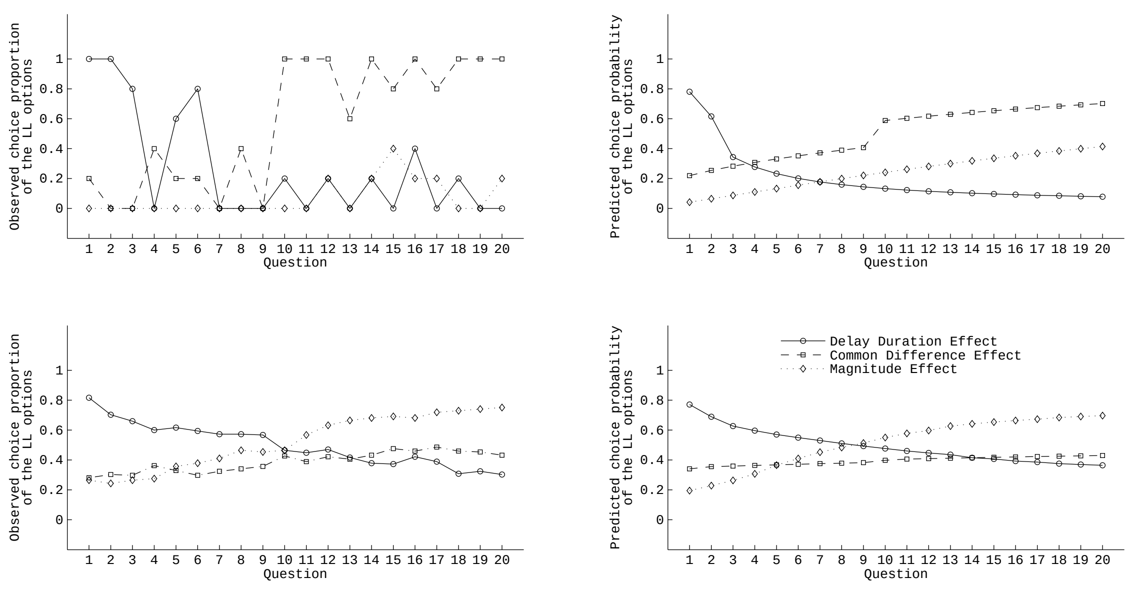

Figure 13. Observed choice proportions and predicted choice probabilities of the LL options for various intertemporal effects in Experiment 3 of Dai and Busemeyer (2014). The left column shows the results of a typical participant (top panel) and the average

results across participants (bottom panel); the right column shows the corresponding predictions of the winning model.

Discussion

An overall model comparison was performed in this section to find a best model among all competing static and dynamic models of intertemporal choice. For this purpose, a small set of candidate models was first figured out using overall AIC and BIC values across all participants to narrow down the number of models for further comparisons. The set of candidate models involved not only the two promising attribute-wise models of intertemporal choice in the literature, i.e., the tradeoff model and the diffusion model selected in Dai and Busemeyer (2014), but also the newly proposed, more general diffusion models with mixed direct differences that might produce even better performance. Therefore, it constitutes a valid foundation for selecting a best model of intertemporal choice when binary choices between gains are of concern.

Various model comparison methods were then applied to the set of candidate models to find the best model, including the comparison of counts of lowest AIC and BIC values across participants, pairwise comparisons using individual AIC and BIC values, a cross-validation test, a Bayesian model comparison, and the comparison of dynamic diffusion models using AIC and BIC when the models were fit to both choice and response time data. In most cases, one or more of the diffusion models performed the best or very close to the best. For example, when only choice data were of concern, the diffusion model with mixed direct differences and no initial bias was the single best model in terms of overall AIC value across participants, overall BIC value across participants, and the result of the cross-validation test. Similarly, when both choice and response time data were taken into consideration, the diffusion model with mixed direct differences and varied z^* was the single best model in terms of overall AIC and BIC values across participants as well as count of lowest AIC values and result of pairwise comparison based on AIC values.

By contrast, the performance of the two static candidate models, i.e., the tradeoff model and the random preference model, were in general inferior to those of the diffusion models. The only exception arose when only choice data were fit and count of lowest AIC or BIC values was used as model selection criterion. In this case, the random preference model was comparable to one or more of the diffusion models. Based on the results of all the model comparisons, it seems appropriate to select the newly proposed diffusion model with mixed direct differences, power transformations of objective value and time allowing for both diminishing and increasing marginal sensitivity, fixed θ^*, varied σ and varied z^* as the best model among all the competing static and dynamic models. It was developed from the diffusion model with single direct differences and no initial bias favored in Dai and Busemeyer (2014), with the diffusion model with mixed direct differences and no initial bias as the intermediate. It was one of the candidate best models when only choice data were considered in model comparisons and the dominating model among the three candidate dynamic models when both choice and response time data were analyzed. Its success was not unexpected since it was developed to account for all the effects and phenomena in intertemporal choice covered in this dissertation. Table 15 lists the relevant effects and phenomena and whether a certain candidate model can qualitatively predict the effects and phenomena. Because each model was built upon direct differences and involved certain form of nonlinear transformations of objective value and time allowing for diminishing marginal sensitivity, all of them could account for the delay duration effect, common difference effect, and magnitude effect in intertemporal choice. By contrast, not all the models can explain the reversal of the common difference effect, the violation of WST, and the relationships between choice proportions and response times revealed in at least the data from a portion of participants. First, because all candidate diffusion models allowed for both convex and concave transformations of objective time but the tradeoff model implied only concave or linear transformation, only the diffusion models could account for the reversal of the common difference effect. Second, both diffusion models with mixed direct differences and the tradeoff model were qualitatively consistent with a potential violation of WST but not the diffusion model with single direct differences nor the random preference model with single direct differences. Third, due to the static nature of the tradeoff model and the random preference model, neither of them could account for the dynamic nature of intertemporal choice demonstrated in the marginal and conditional relationships between choice proportions and response times. Finally, only the winning diffusion model was able to explain both marginal and conditional level relationships because it assumes not only the dynamic structure of diffusion models in general but also an initial bias that has the same sign as the d parameter and thus could predict shorter response times for more frequently chosen option within a pair.

General Discussion

Intertemporal choice has long been investigated with the delay discounting paradigm which assumes a deterministic, static, and alternative-wise perspective on the topic. In this dissertation, I reported the results of three experiments aimed at advocating a fundamentally different approach to intertemporal choice, i.e., the probabilistic, dynamic, and attribute-wise approach. One of the experiments was previously reported in the literature (i.e., Experiment 3 of Dai and Busemeyer [2014]) and the other two were new experiments investigating other aspects of intertemporal choice that should be incorporated into a comprehensive model. Overall, the results of all the experiments suggested that the new perspective on intertemporal choice provided a better description of the empirical data than the traditional perspective.

Table 15

Effects and Phenomena in Intertemporal Choice Explained by the Candidate Models

	Model				
	1	2	3	4	5
Transformations of objective value and time	Power with exponents no larger than 2	Power with exponents no larger than 2	Logarithm	Power with exponents no larger than 2	Power with exponents no larger than 2
Core theory	Attribute-wise with mixed direct differences	Attribute-wise with mixed direct differences	Attribute-wise with single direct differences	Attribute-wise with single direct differences	Attribute-wise with single direct differences
Stochastic specification	Diffusion model with varied σ, fixed θ^* and no initial bias	Diffusion model with varied σ, fixed θ^* and varied z^*	Tradeoff model with ratio rule	Diffusion model with varied σ, fixed θ^* and no initial bias	Random preference model
Effects and Phenomena					
Delay duration effect	√	√	√	√	√
Common difference effect	√	√	√	√	√
Reversal of common difference effect	√	√		√	√
Magnitude effect	√	√	√	√	√
Violation of WST	√	√	√		
marginal relationship between choice proportions and response times	√	√		√	
conditional relationship between choice proportions and response times		√			

Probabilistic, Dynamic, and Attribute-wise Nature of Intertemporal Choice

As mentioned in the introduction section, research on intertemporal choice was initiated by economists such as John Rae and William S. Jevons and had been dominated by economic thoughts for at least a century. As a result, a majority of studies on intertemporal choice were heavily impacted by the DU model (Samuelson, 1937), a standard model of intertemporal choice among economists. Although researchers from various disciplines have attempted to modify the model by incorporating psychological insights into it, most of them take the deterministic, static, and alternative-wise perspective of the DU model as granted. Consequently, much effort has been put to improve the form of the discount function to account for more variability in the observed data of supposedly indifferent pairs of intertemporal options. Recently, however, Scholten and Read (2010) proposed a tradeoff model of intertemporal choice assuming that people process intertemporal choice by weighing the direct difference in money attribute against the direct difference in delay attribute. The authors showed that the tradeoff model can account for a number of anomalies in intertemporal choice that are beyond the means of any alternative-wise models. They also tried to extend the deterministic version of the tradeoff model by using a ratio rule based on the two direct differences to determine the odds of choosing one option against the other. In this way, the model becomes probabilistic and thus could be used to explain choice variability in empirical data. This and following work by Scholten, Read, and colleagues (e.g., Scholten, Read, & Sanborn, 2014) marked an important shift in intertemporal choice literature which imposed a substantial challenge to the traditional approach.

Inspired by the work of Scholten and colleagues, Dai and Busemeyer (2014) ran a series of three experiments on intertemporal choice to examine its probabilistic, dynamic, and attribute-wise nature. These studies were designed around the delay duration effect, the magnitude effect, and the common difference effect in intertemporal choice. It turned out that the probabilistic and dynamic properties of intertemporal choice were revealed in the observed data, while the attribute-wise choice strategy was suggested by the result of an extensive model comparison with more than 90 probabilistic models. This dissertation extends the previous line of research by exploring people's responses to choice questions with immediate SS options, examining the transitivity of intertemporal preference, and performing an even more extensive model comparison with 379 models. The results in general were consistent with previous ones and thus provided converging evidence for the probabilistic, dynamic, and attribute-wise perspective on intertemporal choice.

It is worth noting that the current line of research differs from that of Scholten and colleagues in that the former focuses on individual data while the latter typically relies on aggregate analysis. Either approach has its pros and cons. On the one hand, analyzing individual data excludes aggregate artifact as a possible explanation of the revealed probabilistic and intransitive nature of intertemporal choice, if any, but requires an implicit assumption of independence among responses to the same questions presented repeatedly, which may or may not hold in the actual data. On the other hand, aggregate analysis is easy to conduct and does not need the assumption of independence but it requires an untenable assumption that all participants share the same choice strategies and the same values on the relevant model parameters. A possible way to improve the validity of relevant studies is to collect individual data with repeatedly presented questions as in the current line of research and impose more experimental control to fulfill the assumption of independence as much as possible. In any case, the results of both lines of research provide converging evidence for the attribute-wise nature of intertemporal choice, and the probabilistic and dynamic properties revealed in the current line of research are robust to a possible violation of the independence assumption. In summary, existing evidence strongly supports the probabilistic, dynamic, and attribute-wise view on intertemporal choice. Consequently, a shift in research paradigm on intertemporal choice is highly recommended.

Time Course of Intertemporal Choice

The empirical results concerning the marginal and conditional relationships between choice proportions and response times suggested that intertemporal choice is a dynamic process which requires different amounts of deliberation time for different pairs of options. To account for these relationships, I developed the winning model, i.e., a diffusion model with mixed direct differences, fixed θ^*, varied σ, and varied z^*. The success of this model suggested that the time course of intertemporal choice might be governed by a particular diffusion process. Specifically, this model suggested that a decision maker shifts his/her attention between the two attributes and at each time point samples the direct difference along either attribute as the evidence for or against the two options. The evidence is accumulated over time until the evidence level of one option reaches a threshold first, at which time a decision is made to choose the very option. The model further assumed that the probability of sampling the difference on a particular attribute at any time is contingent on the attention weight on the attribute. As a result, the values of the drift and diffusion parameters of the diffusion process are determined by the specific attribute values involved in each pair of options and the attention weight on each attribute. Finally, it was assumed by the model that there is an initial bias toward choosing the option with a positive mean drift rate. With the evidence accumulation and attention shifting mechanisms in force, the decision maker's preference level of each option changes over time until he/she decides to choose the option that reaches the threshold first. Note that we can also interpret the winning model as if there is no initial bias but the threshold for choosing the option with a positive mean drift rate is lower than that for choosing the other option, which should have a negative mean drift rate for sure. One possible explanation for this mechanism is that people form their initial feeling in favor of the option with a positive mean drift rate in the reading phase when they sample each attribute value before accumulating direct differences as evidence. Based on the initial feeling, they assign different thresholds for the two responses. Alternatively, the initial bias might be the consequence of some learning process that took place when each choice pair was presented repeatedly. Future studies can use a design in which each question is presented only once to examine the alternative possibility.

The time course assumed by the winning model can account for the marginal and conditional relationships between choice proportions and response times as follows. First, according to the winning model, the higher the ratio between the drift and diffusion parameters is in absolute magnitude, the more extreme choice probabilities (and thus observed choice proportions on average) will be and the less time it will require to reach the threshold for a decision. This suggests that, with a range of ratios between the drift and diffusion parameters from different pairs of options, the model can predict the marginal relationship between choice proportions and response times. Second, the model assumes that an option with a positive mean drift rate also enjoys an initial bias toward its choice. This grants the very option not only a predicted choice probability higher than .5 but also a shorter mean response time. Similarly, the other option with a negative mean drift rate is predicted to have a choice probability below .5 as well as a longer mean response time. Stated otherwise, the option chosen more frequently in each pair should also have a shorter mean response time according to the winning model. This is the actual mechanism suggested by the winning model for the observed conditional relationship between choice proportions and response times.

Direct Differences in Subjective Value and Time versus Subjective Direct Differences in (Objective) Value and Time

Another implication of the winning model is that both direct differences in subjective value and time and subjective perceptions of direct differences in (objective) value and time play a role in intertemporal choice. The former is necessary for explaining the common difference effect (and its reversal in the empirical data from Experiment 1) and the latter is necessary for explaining a potential violation of WST in intertemporal choice. The results of quantitative model comparisons also suggested that the inclusion of the latter contributed considerably to model performance, especially when both choice and response time data were taken into consideration. This leads to an important issue of how people process the difference along each attribute to make an intertemporal decision. Both the tradeoff model and the diffusion model with single direct differences favored in Dai and Busemeyer (2014) assume that direct differences are drawn after each objective attribute value is evaluated and assigned a subjective strength, while the current best model suggests that both types of direct differences might be considered and accumulated in the deliberation process. A close look at the distribution of the estimated values of parameter k in the winning model, however, revealed that 120 out of the 138 participants had an extreme value on the parameter (i.e, below .05 or above .95), suggesting that a majority of participants considered either the direct differences in subjective value and time or the subjective direct differences in (objective) value and time, but not the both. Specifically, 65 participants had a k value above .95, indicating considering only direct differences in subjective value and time, and 55 participants had a k value below .05, suggesting sole attention to subjective perceptions of direct differences in (objective) value and time. For the remaining 18 participants, the estimated values on k parameter were in favor of a mixed approach to considering the direct differences in intertemporal choice. In summary, to maintain a reasonably good performance of the winning model for each participant, it seems necessary to include both types of direct differences in the model.

Diminishing versus Increasing Marginal Sensitivity to Monetary Rewards

The potential convex utility function suggested by the winning model may appear alarming given the widely accepted notion of diminishing marginal utility. However, anecdotal and experimental evidence in favor of the concept of diminishing marginal utility usually involves relatively large amounts of rewards [e.g., thousands of dollars as used in Kahneman & Tversky, 1979] and/or certain form of risk [e.g., in the famous St. Petersburg Paradox presented in Bernoulli (1954)]. To the contrary, the reward amounts participants encountered in the current studies were relatively small ones so that the actual payment could be associated with the encountered values to make the task more incentive. Furthermore, there was not an obvious risky component associated with the rewards in the current experimental setting. Therefore, it should not be too surprising that a model allowing for convex utility functions won the competition against other models consistent with the notion of diminishing marginal utility. More importantly, when rewards are relatively small as in the current studies, another mechanism may kick in that actually requires convex utility functions, at least around certain context-dependent points. Specifically, participants may pick a couple of highest rewards they encountered in the studies as their goals and therefore assign disproportional utilities to these rewards. As a result, the best fitting value for the exponent in the power utility function should be greater than one to capture the impact of goal-setting in intertemporal choice. This notion was supported by my previous study on this issue (Dai, 2009) and consistent with the tri-reference point theory of decision making under risk (Wang & Johnson, 2012). In summary, it is possible that rewards in intertemporal choice and rewards in risky choice have different utility functions, at least for small amounts of rewards as in current line of research. It is worth mentioning that the winning model also allows for the traditional form of power utility function that implies diminishing marginal utility. Therefore, the issue concerning the appropriate value for the exponent could be resolved by fitting the model to actual data.

Extension of the Winning Model to Binary Intertemporal Choice Between Losses

Intertemporal choice might involve not only positive outcomes but also negative ones and people may behave differently in these two situations. For example, previous studies (e.g., Thaler, 1981) have shown that the discount factor for losses tend to be higher (i.e., less discounting) than that for gains, leading to the so-called gain-loss asymmetry in intertemporal choice. As a result, a general model of intertemporal choice should be applicable to the domain of losses and more importantly it should be able to explain the gain-loss asymmetry. There is actually no specific constraint on the winning model which prevents it from being applied to negative outcomes. The power utility function can be applied to negative outcomes, with or without invoking the assumption of loss aversion. And all the other components of the model can be simply implemented in the same way as for gains. The gain-loss asymmetry can be explained by assuming that losses attract more attention than gains, which is not unnatural. As a result, the direct difference on the money attribute has a higher impact with negative outcomes as opposed to positive ones. Consequently, when signs of both payoffs are reversed from positive to negative, people would shift their preference to the SS option because it provides a better outcome than the LL option. In other words, if one is indifferent between receiving $10 now and receiving $20 in a month, then he or she should prefer losing $10 now as opposed to losing 20 dollars in a month. This is exactly what the gain-loss asymmetry in intertemporal choice suggests. In summary, the current winning model can be conveniently extended to the area of intertemporal choice between losses and it is capable of explaining the relevant finding in the literature even without invoking the notion of loss aversion. It is worth noting that the loss aversion assumption only strengthens the prediction of the winning model on the gain-loss asymmetry and thus will not incur a problem with the current model.

Individual Differences in Intertemporal Choice and Its Implication on Model Selection and Application

Finally, it should be mentioned that the best model did not perform the best for each individual and there were clearly individual differences in intertemporal choice in terms of both the estimated parameter values when a specific model was fit to individual data and the best model for each individual. For example, when both choice and response time data were considered, the overall best model outperformed the other two candidate diffusion models for 88 out of the 138 participants with regard to AIC value and 69 out of the 138 participants with regard to BIC value. The individual data from remaining participants were actually best described by the diffusion model with mixed direct differences and no initial bias. Furthermore, within the participants whose data were best described by the overall best model, there were certainly individual differences in parameter estimates, suggesting individual differences in the various components of the overall best model. When only choice data were considered, individual differences appeared even more prominent.

The existence of individual differences with regard to the best model for each participant gives rise to an important issue in model selection and application: Shall we fit an overall best model, if any, to individual data of all participants, or use different best models for different participants? On the one hand, it is unlikely that choice behavior of each participant is governed by exactly the same mechanisms suggested by the overall best model. Consequently, the conclusion based on parameter estimates from the single overall best model might be misleading. On the other hand, using different models for different participants will make it difficult to compare their parameter estimates and make meaningful conclusions. Furthermore, although different model selection methods may suggest different best models for different participants, it does not necessarily mean that participants actually process intertemporal choice in different ways. The probabilistic nature of intertemporal choice makes it possible that data generated by one model are best fit by a different but similar model. Since the current model repertoire was generated in a factorial way, different models tended to share common components to some extent. Consequently, the aforementioned situation might well occur in the current model comparison procedure. Finally, the three dynamic models in the candidate set were actually nested models with the overall best model as the most general one. Therefore, we could simply use the overall best model and let parameter estimates tell whether a reduced model is more appropriate for describing individual data. In summary, it seems a reasonable expediency at this stage to apply the overall best model in new studies to detect potential individual differences underlying intertemporal choice.

Conclusions

This dissertation advances my previous empirical and modeling work on the probabilistic, dynamic, and attribute-wise nature of intertemporal choice. Two new experiments with different types of intertemporal choice questions and an even more extensive comparison of competing static and dynamic models were conducted to further examine the relevant properties and look for a more comprehensive cognitive model of intertemporal choice. It turned out that the probabilistic, dynamic, and attribute-wise nature of intertemporal choice appeared no matter whether the SS options occurred immediately (as in Experiment 1) or with some delays as the LL options [as in Experiment 2 of this dissertation and Experiment 3 in Dai and Busemeyer (2014)]. In addition, most participants showed transitive intertemporal preferences in terms of weak stochastic transitivity. The overall best model of the current model comparison is a generalization of the diffusion model with direct differences advocated in Dai and Busemeyer. This model can account for all the effects and phenomena examined in this dissertation, including the delay duration effect, the common difference effect (and its reversal), the magnitude effect, the potential intransitivity of intertemporal preference, and the marginal and conditional relationships between choice proportions and response times. Furthermore, this model can be conveniently extended to intertemporal choice between losses and account for the relevant gain-loss asymmetry. Consequently, it deserves a recommendation as a better replacement for the existing models of intertemporal choice which assume a deterministic, static, and alternative-wise perspective on the topic.

References

Ainslie, G. (1974). Impulse control in pigeons. *Journal of the Experimental Analysis of Behavior*, *21*, 485-489.

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin*, *82*, 463-496.

Ainslie, G., & Herrnstein, R. (1981). Preference reversal and delayed reinforcement. *Animal Learning and Behavior*, *9*, 476-482.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716-723.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1963). Stochastic models of choice behavior. *Behavioral Science*, *8*, 41-55.

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk. *Econometrica*, *22*, 23-36.

Böhm-Bawerk, E. V. (1889). *Capital and interest.* Libertarian Press.

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without tradeoffs. *Psychological Review*, *113*, 409-432.

Broomell, S. B., Budescu, D. V., & Por, H.-H. (2011). Pair-wise comparisons of multiple models. *Judgment and Decision Making*, *6*, 821-831.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153-178.

Busemeyer, J. R., & Diederich, A. (2009). *Cognitive modeling.* Sage Publications.

Busemeyer, J. R., & Townsend, J. T. (1992). Fundamental derivations from decision field theory. *Mathematical Social Sciences*, *23*, 255-282.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*, 432–432.

Cavagnaro, D. R., & Davis-Stober, C. P. (in press). Transitive in our preferences but transitive in different ways: An analysis of choice variability. *Decision*.

Chapman, G., & Winquist, J. (1998). The magnitude effect: Temporal discount rates and restaurant tips. *Psychological Bulletin & Review*, *5*, 119-123.

Christensen-Szalanski, J. (1984). Discount functions and the measurement of patients' values. *Medical Decision Making*, *4*, 47-58.

Dai, J. (2009). *Delay discounting and social discouting: Two empircal studies* (Unpublished master's thesis). University of South Dakota.

Dai, J., & Busemeyer, J. R. (2014). A probabilistic, dynamic, and attribute-wise model of intertemporal choice. *Journal of Experimental Psychology: General*, *Online first publication*.

Davidson, D., & Marschak, J. (1959). Measurements: Definitions and theories. In C. W. Churchman & P. Ratoosh (Eds.), (p. 274). John Wiley & Sons, Inc.

Diederich, A. (2003). Mdft account of decision making under time pressure. *Psychological Bulletin & Review*, *10*, 157-166.

Fisher, I. (1930). *The theory of interest.* New York: Macmillan.

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, *40*(2), 351-401.

González-Vallejo, C. (2002). Making trade-offs: A probabilistic and context-sensitive model of choice behavior. *Psychological Review*, *109*, 137-155.

González-Vallejo, C., & Reid, A. A. (2006). Quantifying persuasion effects on choice with the decision threshold of the stochastic choice model. *Organizational Behavior and Human Decision Processes*, *100*, 250-267.

Green, L., Fisher, B., Perlow, S., & Sherman, L. (1981). Preference reversal and self-control: Choice as a function of reward amount and delay. *Behavioral Analysis Letters*, *1*, 43-51.

Green, L., Fry, A. F., & Myerson, J. (1994). Discounting of delayed rewards: A life-span comparison. *Psychological Science*, *5*, 33-36.

Green, L., Myerson, J., & McFadden, E. (1997). Rate of temporal discounting decreases with amount of reward. *Memory and Cognition*, *25*, 715-723.

Grether, D., & Plott, C. (1979). Economic theory of choice and the preference reversal phenomenon. *The American Economic Review*, *69*, 623-638.

Harless, D., & Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, *62*, 1251-1289.

Jevons, H. S. (1905). *Essays on economics*. Macmillan.

Jevons, W. S. (1888). *The theory of political economy*. Macmillan.

Johnson, J. G., & Busemeyer, J. R. (2005). A dynamic, stochastic, computational model of preference reversal. *Psychological Review*, *112*, 841-861.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263-291.

Kirby, K. N., & Herrnstein, R. J. (1995). Preference rreversals due to myopic discounting of delayed reward. *Psychological Science*, *6*, 83-89.

Kirby, K. N., & Marakoviæ, N. N. (1996). Delay-discounting probabilistic rewards: Rates decrease as amounts increase. *Psychological Bulletin & Review*, *3*, 100-104.

Klugkist, I., & Hoijtink, H. (2007). The bayes factor for inequality and about equality constrained models. *Computational Statistics and Data Analysis*, *51*(12), 6367-6379.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*, 1292-1298.

Kruschke, J. K. (2010). *Doing bayesian data analysis*. Academic Press: Elsevier.

Loewenstein, G., & Prelec, D. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *The Quarterly Journal of Economics*, *107*, 573-597.

Loomes, G., & Sugden, R. (1995). Incorporating a stochastic element into decision theories. *European Economic Review*, *39*, 641-648.

Luhmann, C. C. (2013). Discounting of delayed rewards is not hyperbolic. *Journal of*

Experimental Psychology: Learning Memory and Cognition, *39*(4), 1274-1279.

Mazur. (1987). Quantitative analysis of behavior: Vol 5. the effect of delay and intervening events on reinforcement value. In M. J. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), (p. 55-73). Mahwah, NJ: Erlbaum.

McKerchar, T. L., Green, L., Myerson, J., Pickford, T. S., Hill, J. D., & Stout, S. C. (2009). A comparison of four models of delay discounting in humans. *Behavioral Processes*, *81*, 256-259.

Millar, A., & Navarick, J. (1984). Self-control and choice in humans: Effects of video game playing as a positive reinforcer. *Learning and Motivation*, *15*, 203-218.

Mosier, C. I. (1951). Problems and designs of cross-validation. *Educational and Psychological Measurement*, *11*, 5-11.

Mosteller, F., & Nogee, P. (1951). An experimental measurement of utility. *Journal of Political Economy*, *59*, 371-404.

Myung, I. J., & Pitt, M. A. (1997). Applying occam's razor in modeling cognition: A bayesian approach. *Psychonomic Bulletin and Review*, *4*(1), 79-95.

Prelec, D., & Loewenstein, G. (1991). Decision making over time and under uncertainty: A common approach. *Management Science*, *37*, 770-786.

Rachlin, H. (2006). Notes on discounting. *Journal of Experimental Analysis of Behavior*, *85*, 425-435.

Rae, J. (1905). *The sociological theory of capital.* New York: Macmillan.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59-108.

Ratcliff, R., & Smith, P. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*, 333-367.

Read, D. (2001). Is time-discounting hyperbolic or subadditive? *Journal of Risk and Uncertainty*, *23*, 5-32.

Read, D., & Roelofsma, P. H. M. P. (2003). Subadditive versus hyperbolic discounting: A comparison of choice and matching. *Organizational Behavior and Human Decision*

Processes, *91*, 140-153.

Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2011). Transitivity of preferences. *Psychological Review*, *118*, 42-56.

Regenwetter, M., & Davis-Stober, C. P. (2012). Behavioral variability of choice versus structural inconsistency of preferences. *Psychological Review*, *119*, 408-416.

Richards, J. B., Zhang, L., Mitchell, S. H., & De Wit, H. (1997). Delay of probability discounting in a model of impulsive behavior: Effect of alcohol. *Journal of Experimental Analysis of Behavior*, *67*, 353-366.

Roe, R., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review*, *108*, 370-392.

Roelofsma, P. H. M. P., & Read, D. (2000). Intransitive intertemporal choice. *Journal of Behavioral Decision Making*, *13*, 161-177.

Samuelson, P. (1937). A note on measurement of utility. *The Review of Economic Studies*, *4*, 155-161.

Scholten, M., & Read, D. (2010). The psychology of intertemporal tradeoffs. *Psychological Review*, *117*, 925-944.

Scholten, M., Read, D., & Sanborn, A. (2014). Weighing outcomes by time or against time? evaluation rules in intertemporal choice. *Cognitive Science*, 1-40.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461-464.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*, 153-181.

Strotz, R. H. (1956). Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies*, *23*(3), 165-180.

Takahashi, T., Oono, H., & Radford, M. H. B. (2008). Psychophysics of time perception and intertemporal choice models. *Physica A*, *387*, 2066-2074.

Thaler, R. (1981). Some empirical evidence on dynamic inconsistency. *Economic Letters*, *8*, 201-207.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*, 273-286.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, *76*(1), 31-48.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297-323.

Tversky, A., & Russo, J. E. (1969). Substitutability and similarity in binary choices. *Journal of Mathematical Psychology*, *6*(1), 1-12.

Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamic models of multialternative choice. *Psychological Review*, *111*, 757-769.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior.* Princeton, NJ: Princeton University Press.

Wang, X. T., & Johnson, J. G. (2012). The tri-reference point theory of decision making under risk. *Journal of Experimental Psychology: General*, *141*(4), 743-756.

Weber, E. U., Johnson, E. J., Milch, K. F., Chang, H., Brodscholl, J. C., & Goldstein, D. G. (2007). Asymmetric discounting in intertemporal choice: A query-theory account. *Psychological Science*, *18*, 516-523.

Wittmann, M., & Paulus, M. P. (2008). Decision making, impulsivity and time perception. *Trends in Cognitive Sciences*, *12*(1), 7-12.

Zauberman, G., Kyu Kim, B., Malkoc, S. A., & Bettman, J. R. (2009). Discounting time and time discounting: Subjective time perception and intertemporal preferences. *Journal of Marketing Research*, *46*, 543-556.

Appendix Prior Distributions Used in the Bayesian Model Comparison

For the Bayesian model comparison, I used an uninformative prior distribution for each model by setting a uniform distribution within a reasonable range for each parameter. Specifically, for attribute-wise diffusion models, I set

$$w \sim uniform(.05, .95), \tag{47}$$

$$\alpha, \beta \sim uniform(.01, 2), \tag{48}$$

$$\theta* \sim uniform(0, 5), \tag{49}$$

$$z* \sim uniform(0, .9 \cdot \theta*), \tag{50}$$

$$k \sim uniform(0, 1). \tag{51}$$

For the tradeoff model, I set

$$\gamma, \tau, \kappa, \alpha, \epsilon \sim uniform(0, 20), \tag{52}$$

$$\ell \sim uniform(1, 21). \tag{53}$$

Finally, for the random preference model, I set

$$\alpha, \beta \sim uniform(.01, 2), \tag{54}$$

$$a, b \sim uniform(1, 50), \tag{55}$$

in which a and b are the shape parameters of the beta distribution for the attention weight parameter. The ranges of parameter values (except that for the k parameter) involved in the attribute-wise diffusion models were derived from the estimated values of the winning model in Dai and Busemeyer (2014), i.e., the diffusion model with single direct difference, fixed θ^*, varied σ, and no initial bias. Because the k parameter was a weight parameter between two types of direct differences, it was natural to constrain it between 0 and 1. For the tradeoff model, the ranges of parameter values were derived from the prior distribution used in a recent article comparing the tradeoff model with other static models of intertemporal choice (Scholten et al., 2014). Specifically, I transformed the informative priors used in that article into uninformative ones but set the ranges of parameter values to cover the main parts of the previous informative priors. In this way, all candidate models were defined on uninformative priors. Finally, for the random preference model, the exponents of power transformations, i.e., α and β, had the same range as those in the attribute-wise diffusion models, and the range of the shape parameters of the beta distribution was set to cover a wide range of distribution shape of the attention weight parameter from a uniform distribution to quite peaked distributions.

Footnotes

[1]In this dissertation only positive outcomes are considered and thus Equation 8 is only for gains. However, the resultant best model can be easily extended to explain intertemporal choice between losses.

[2]In this dissertation I assume a two-alternative forced choice paradigm and probability zero of having no preference between the two options. See (Regenwetter & Davis-Stober, 2012)Regenwetter and Davis-Stober (2012) for a more general form of random utility models.

[3]With a continuous distribution on the principal parameter, the probability of having no preference between any pair of options, i.e., $U_A = U_B$, equals zero.

[4]Data of two participants were not involved in the related-samples t test because they appeared deterministic after data pruning and thus did not provide valid numbers for comparison. The same was true for the following conditional analysis.

[5]Data of four participants were not involved in the related samples t-test because each choice proportion was in the extreme range (i.e., below .2 or above .8) and thus did not provide valid numbers for comparison.

Junyi Dai

Indiana University, Bloomington

Department of Brain and Psychological Sciences

Cognitive Psychology Program

Judgment and Decision Making Lab

Phone: 812-856-4678

Email: junydai@indiana.edu

Address: 1101 E. 10th Street,

Bloomington, IN 47405, USA

Research Interests

Judgment and Decision Making, Cognitive Modeling, Bayesian Data Analysis and Model Comparison

Education

Indiana University, Bloomington, Indiana USA

Ph.D., Psychology, July 2014

Advisor: Dr. Jerome R. Busemeyer

Indiana University, Bloomington, Indiana USA

M.S., Applied Statistics, January 2014

University of South Dakota, Vermillion, South Dakota USA

M.A., Psychology, August 2009

East China Normal University, Shanghai, China

M.S., Psychology, July 2005

Fudan University, Shanghai, China

B.S., Physics, July 2002

Publications

Dai, J., & Busemeyer, J. R. (in press). A Probabilistic, Dynamic, and Attribute-wise Model of Intertemporal Choice. *Journal of Experimental Psychology: General*.

Presentations

Conference Presentations

An extensive comparison of probabilistic models on intertemporal choice. The 46th annual meeting of the Society of Mathematical Psychology, Germany, 2013.

The probabilistic nature of intertemporal choice. The 2012 Society of Judgment and Decision Making annual conference, USA, 2012.

A comparison of stochastic models of intertemporal choice. The 45th annual meeting of the Society of Mathematical Psychology, USA, 2012.

Towards a dynamic, probabilistic, and attribute-wise model of intertemporal choice. FUR XV international conference, USA, 2012.

The probabilistic nature of intertemporal choice. The 2nd Midwest Cognitive Science Conference, USA, 2012

Simulant modeling and experimental study of bargaining game. The 10th Chinese Academic Conference of Psychology. China, 2006.

Poster Presentations

A random utility model on delay discounting and its application to disinhibitory psychopathology via maximum likelihood and Bayesian parameter estimation. The 2nd Midwest Cognitive Science Conference, USA, 2012.

Towards a dynamic stochastic model of intertemporal choice, The 1st Midwest Cognitive Science Conference, USA, 2011.

Awards and Honors

William K. Estes Summer Research Award, Department of Brain and Psychological Sciences at Indiana University, 2013

Society for Mathematical Psychology Student Travel Award, 2013

Commendation on Qualifying Exam Performance, Department of Brain and Psychological Sciences at Indiana University, 2012

Municipal Excellent Degree Thesis Award, Shanghai, 2006

Teaching Experience (undergraduate courses)

Statistical Techniques, Indiana University, Bloomington (2013)

Statistics in Psychological Research, University of South Dakota (2009, 2010)

Professional Service

Ad Hoc Reviewer: Cognition; Decision; Journal of Experimental Psychology: General; Judgment and Decision Making; Journal of Mathematical Psychology; Psychological Methods; Psychological Review

Professional Affiliation

Society for Mathematical Psychology

Society for Judgment and Decision Making

Programming Skills

Matlab

Visual Basic

R

STAN

WinBUGS

Statistical Analysis Skills

Bayesian data analysis and model comparison

General linear models

Generalized linear models

Structural equation modeling

Categorical data analysis

Professional Training

Graduate Summer School, Probabilistic (Bayesian) models of cognition, UCLA, 2011